Seven Pines Symposium XV, May 18-22, 2011

On the origin of genetic code: Was it a frozen accident?

Sergei Rodin

Theoretical Biology, Department of Molecular and Cellular Biology

Beckman Research Institute, City of Hope,

Duarte, CA., USA

City of Hope®

mRNA (gene)

5'---UUC GCC GUC--->3'

N---Phe Ala Val--->C

protein

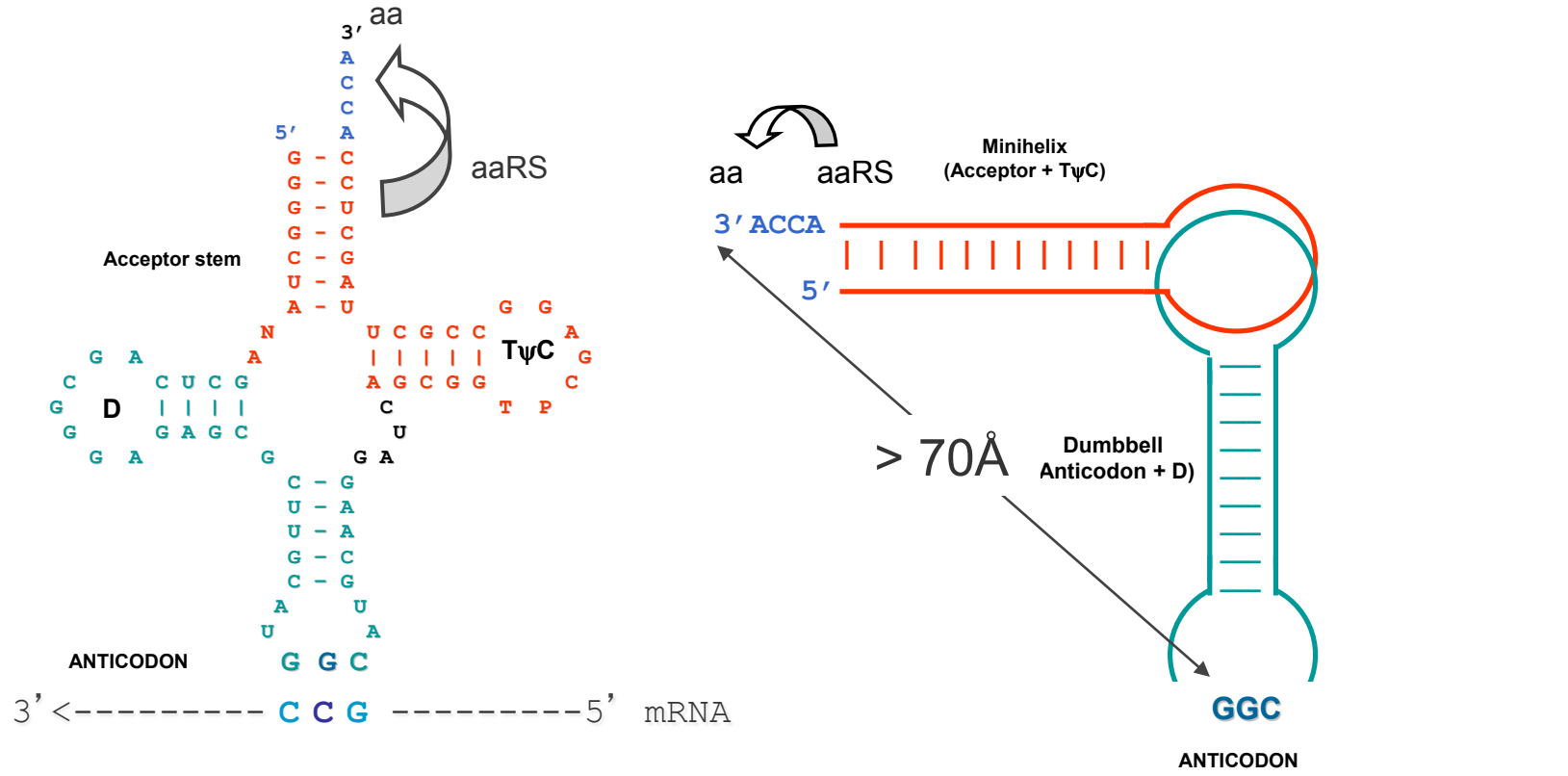
F. Crick (1955): A special class of bifunctional "adaptors" must have existed; otherwise, it would be difficult for the hydrophilic bases in mRNA to form hydrophobic pockets specifically accommodating at least aliphatic and aromatic amino acids

Discovery of a transfer RNA confirmed his insight but brought about numerous ORIGIN problems.

5'	U		C		A		G		3'
	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
U	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	*Stop*	UGA	*Stop*	A
	UUG	Leu	UCG	Ser	UAG	*Stop*	UGG	Trp	G
	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
C	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
A	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
G	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

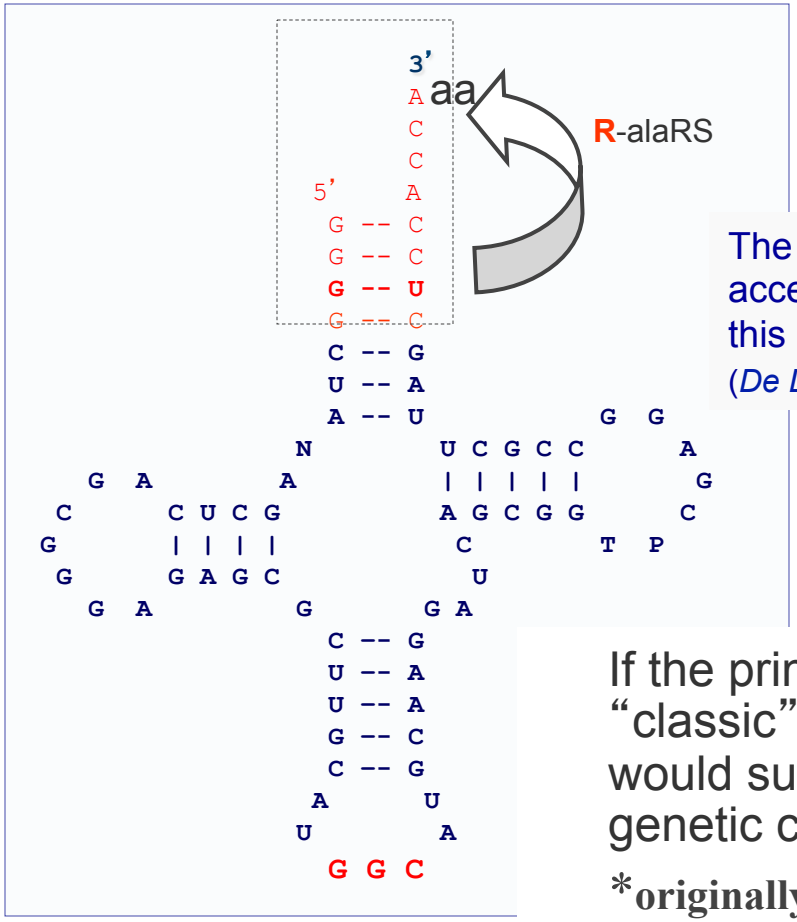
Origin of the genetic code: the main paradox

tRNA 2D clover leaf **tRNA 3D L-shape**



In the 2D cloverleaf and L-shaped 3D structure of tRNAs, the anticodon and 3’ terminal site of amino acid attachment are separated by a maximum distance

Many of present-day tRNAs truncated to the acceptor micro-helix contain sufficient information to be charged by the correct amino acid. Reciprocal truncations of aminoacyl-tRNA synthetases (aaRS), such that in extreme cases the reduced enzyme cannot even physically extend to cover the anticodon, did not change the specificity of aminoacylation as well (*Schimmel et al., 1988, 1993…*)



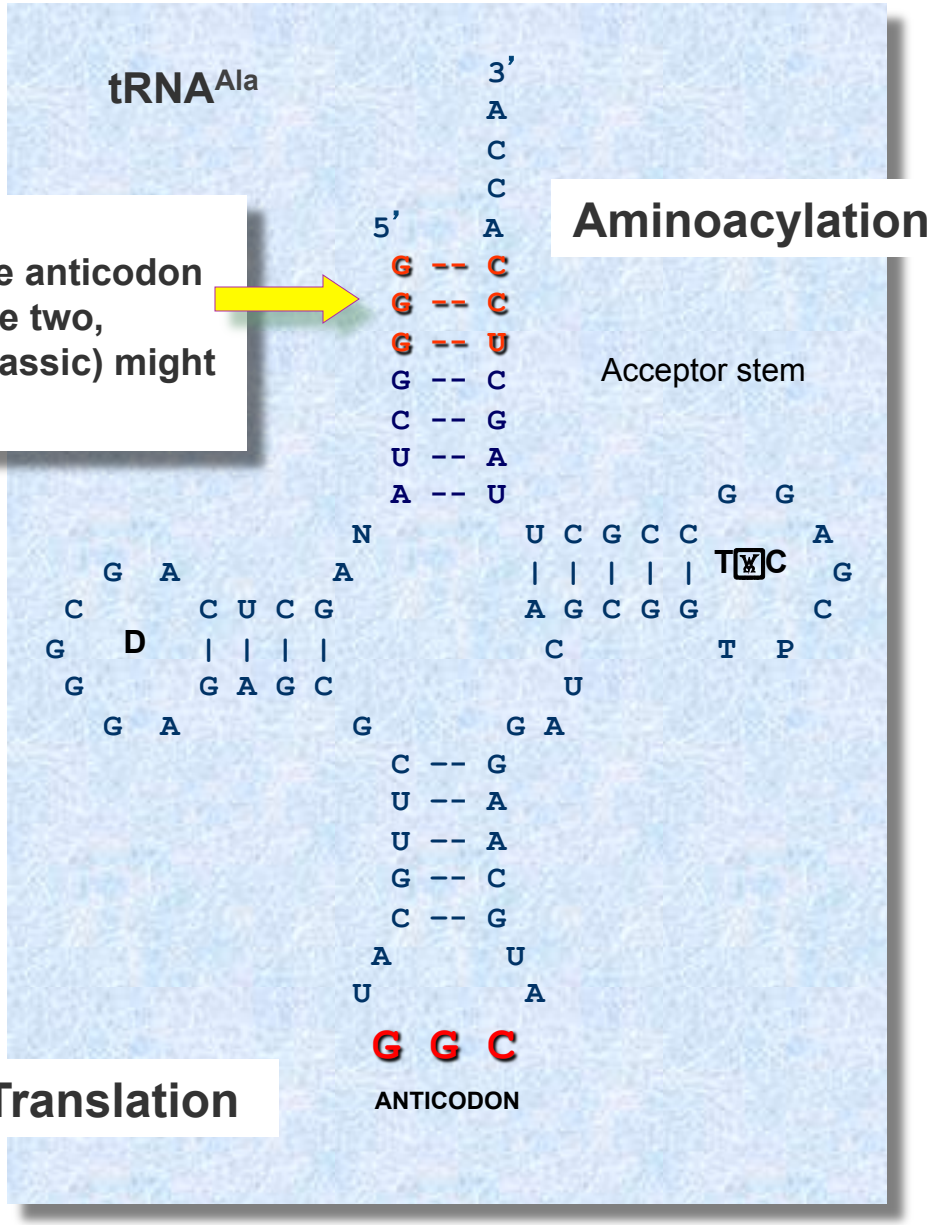
The idea of the "second", operational, code in the acceptor helix. As more essential for aminoacylation, this code could be older than the classic genetic code (*De Duve, 1988; Schimmel et al., 1993*)..

If the primordial operational* and present-day "classic" codes never had anything in common, it would support the hypothesis *(Crick, 1968)* that the genetic code originated as a frozen accident…

***originally implemented by ribozymic precursors of aaRSs**

PARADOX OF TWO CODES

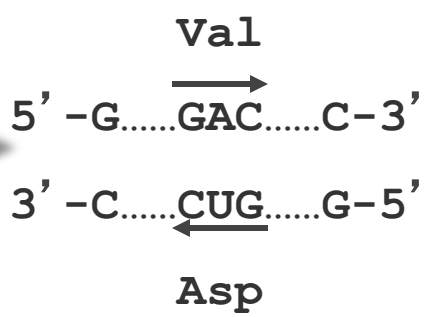
The only reasonable solution – a duplication of the anticodon within the same tRNA molecule meaning that these two, presently very different, codes (operational and classic) might have had a common ancestor



DUPLICATION?

"Disappointingly", all attempts to decipher the acceptor code in terms of simple "words" resembling the anticodons (or codons) have failed…

Ancestral tRNAs with complementary anticodons turned out to be complementary in the 2nd base pair of the acceptor as well (*Rodin, Rodin & Ohno, 1996; updated in: Rodin, Szathmáry & Rodin, 2009)*



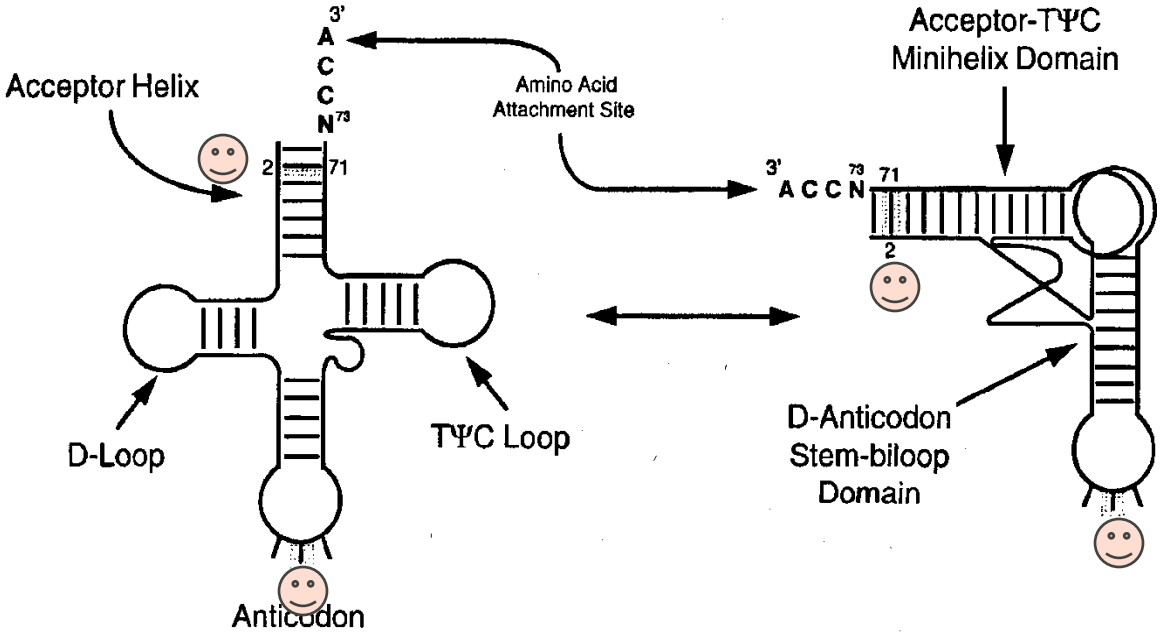


FIG. 1. Schematic diagram of tRNA cloverleaf (*Left*) and of L-shaped three-dimensional structure comprised of two domains (*Right*). The 2–71 bp in the acceptor helix and the second base of the anticodon are shaded. (Illustration provided by Dr. Barry Henderson.)

Dual complementarity is the strong (however indirect) evidence of the historical link between the operational and classic codes.

↓

Primordial codon-anticodon pairs might have been placed in the closest vicinity to the CCA-3' end.

" It is clear that at some early stage in the evolution of life the direct association of amino acids with polynucleotides, which was later to evolve into the genetic code, must have begun."

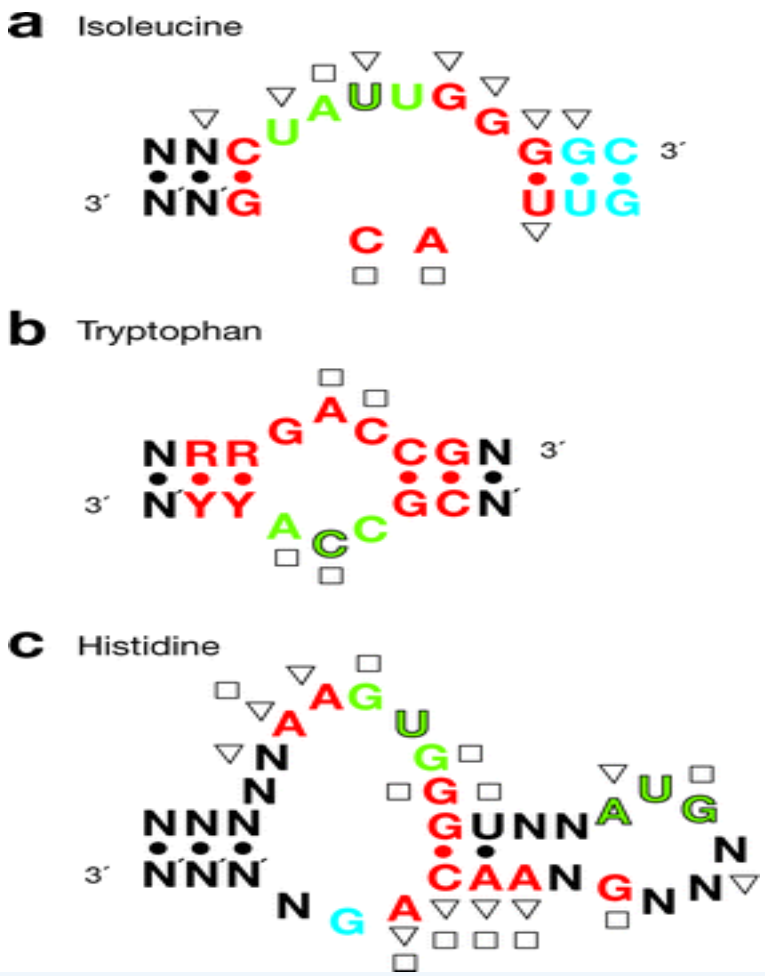
Orgel 1968

Testing the “ key-lock vs. frozen accident” dilemma by a closer analysis of:

1 The updated library of aa-binding sites in (*in vitro* evolved) RNAs that might refer to the most early stereochemical era in the history of the genetic code (*Yarus et al., 2009, 2010*), and

2 The genetic code structure *per se*, as well as its adaptors (tRNAs) and “implementers” (aaRSs).

Part 1

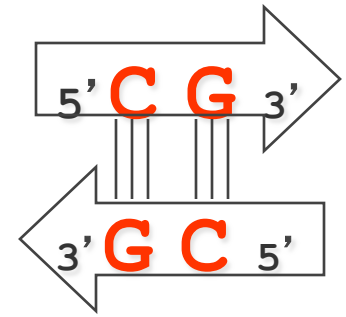


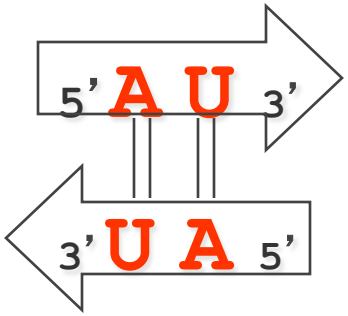
By now, aa-binding RNA aptamers have been successfully "selexed" for nine amino acids. *(Yarus et al., 2009).* Their binding sites are of the three types:

- **Sites in which cognate codon and anticodon are both significantly over-represented: ARG, ILE, and (with narrowly missing significance) TYR**
- Sites in which only cognate anticodons are found in significant excess: HIS, PHE, TRP, and presumably LEU and (?)VAL
- Sites in which neither anticodons nor codons significantly dominate: GLN

The updated compilation already has 468 RNA aptamers and yet, remarkably, the "codons only" group remains empty!

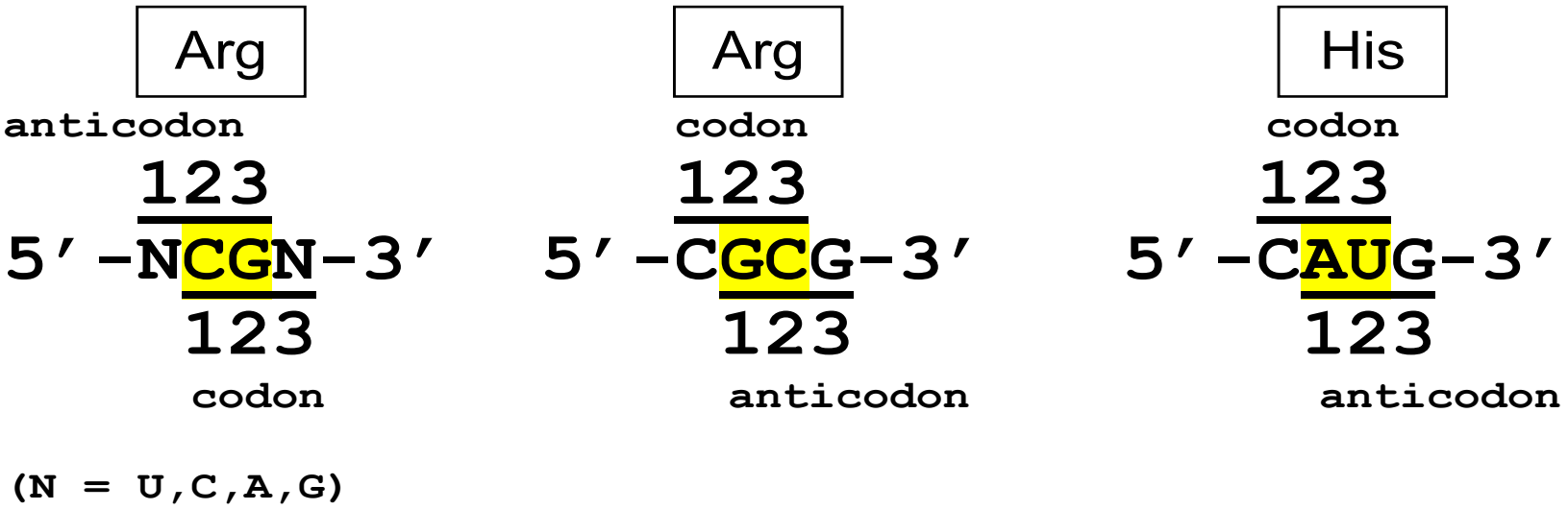
Amino acid aptamers that are arguably the simplest possible specific, functional RNAs. All colored nucleotides are conserved. Red indicates 90% to complete conservation; blue, 60% to 80% conservation; and green, conserved coding triplets. **Green nucleotides are totally conserved**. Squares indicate ligand-protected or enhanced chemical reactivities; triangles show chemical modification-interference with ligand binding (*Yarus et al., 2005)*





5’	U		C		A		G		3’
	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
U	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	*Stop*	UGA	*Stop*	A
	UUG	Leu	UCG	Ser	UAG	*Stop*	UGG	Trp	G
	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
C	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
A	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
G	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

Self-complementarity of the CG dinucleotide increases probability of finding codon in an Arg-binding site, if anticodon is already there (and *vice versa*.)



- The same is the case of four other amino acids that have palindrome-dinucleotide containing codons: AUN (Ile, Met), UAY (Tyr), GCN (Ala).
- In the middle: a particular codon of arginine, CGC, which contains a CG palindrome at 1-2 positions and simultaneously a GC palindrome at 2-3 positions. Accordingly, if the next nt is G, one gets the anticodon with the same palindrome GC at 1-2 positions.
- In contrast, histidine's codon CAU has AU palindrome at 2-3 positions **only**, hence its anticodon (AUG) appears with the same AU at 1-2 positions.

at first two positions…

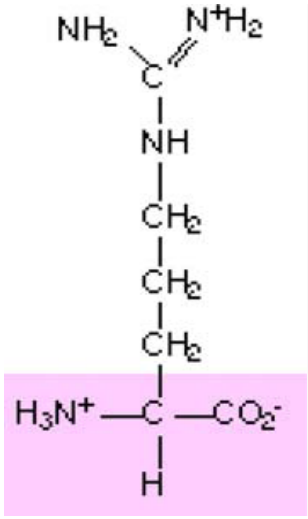
CG-, GC-, UA- and AU-containing codons

- In the complete code, a number of codons with any of such palindromes at 1-2 positions is equal to that with the same palindrome at 2-3 positions.

- **There are no reasons whatsoever for *a priori* belief that the procedure of RNA aptamers selection *per se* --- the selection focused on stereo-specific binding of amino acids to particular RNA sequences --- has anything to do with translation, charging tRNAs with cognate amino acids, wobbling interface between codon and anticodon at their 3rd and 1st bases, etc.**

- **One would think that these two perfectly symmetric cases, (1-2)/(2-3) and (2-3)/(1-2) should be equally represented in aa-binding sites of selexed RNA aptamers**.

  However, in reality…



Arginine

codons	anticodons
CGU	ACG
CGC	GCG
CGA	UCG
CGG	CCG

Palindrome **CG**
codon(1-2)/anticodon(2-3)

Codon	123	123	123	123	123	123
	CGC ->	NCGC:	UCGC	CCGC	ACGC	GCGC
Anticodon		123	123	123	123	123
Anticodon	123	123	123	123	123	123
	GCG ->	GCGN:	GCGA	GCGG	GCGU	GCGC
Codon		123	123	123	123	123

5’ -G**CG**G-3’ **Overrepresented**
(anticodon drives)

VS.

Palindrome **GC**
codon(2-3)/anticodon(1-2)

codon	123	123	123	123	123	123
	CGC ->	CGCN:	CGCU*	CGCC*	CGCA*	CGCG
Anticodon		123	123	123	123	123
Anticodon	123	123	123	123	123	123
	GCG ->	NGCG:	*AGCG	*GGCG	*UGCG	CGCG
Codon		123	123	123	123	123

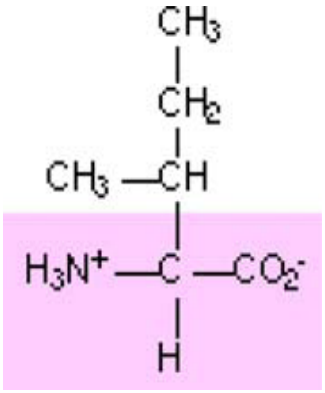
5’ -C**GC**G-3’ : **NONE!**

Marked by asterisk are the cases when codon (anticodon) represents a different amino acid

Arg # 17:

Arg-binding Arg-binding

ucgcggugggguucggugCAAGGA**GCGG**uuuaaaugcCAGGUAGG**UCG**gcaccgau



Isoleucine :

Codons	Anticodons
AUU	AAU
AUC	GAU
AUA	UAU

Palindrome **AU**
codon(1-2)/anticodon(2-3)

```
Codon          123      123     123     123     123     123
               AUA -> NAUA:   UAUA   *CAUA   AAUA   GAUA
Anticodon              123     123     123     123     123

Anticodon      123      123     123     123     123     123
               UAU -> UAUN:   UAUA   UAUG*  UAUU   UAUC
Codon                  123     123     123     123     123
```

VS.

Palindrome **UA**
codon(2-3)/anticodon(1-2)

```
codon          123      123     123     123     123     123
               AUA -> AUAN:   AUAU   AUAC*  AUAA*  AUAG*
Anticodon              123     123     123     123     123

Anticodon      123      123     123     123     123     123
               UAU -> NUAU:   AUAU   *GUAU  *UUAU  *CUAU
Codon                  123     123     123     123     123
```

5’-UAUU-3’
the tetraplet over-represented in Ile-binding sites of selexed RNA aptamers **(184 of 185!)**
Marked by asterisk are the cases when codon (anticodon) represents a different amino acid

5’-AUAU-3’ : NONE!

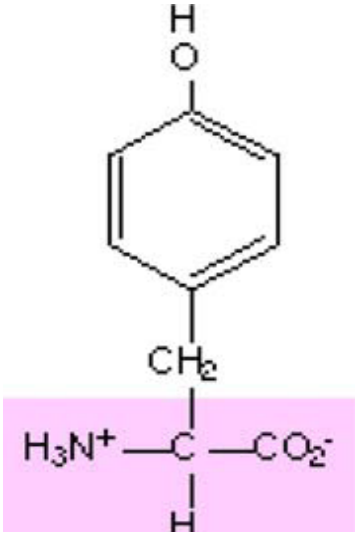
Ile #250 …cgcCUAUUGGGGCcugaugcgcguugggcaagu auac cuugac…agguuACG…

Ile #235b…caaCUAUUGGGUgacuuac auau ugcuaggacagcucagucaaa…gucauACG…

Tyrosine:

(complementary partner of Ile)

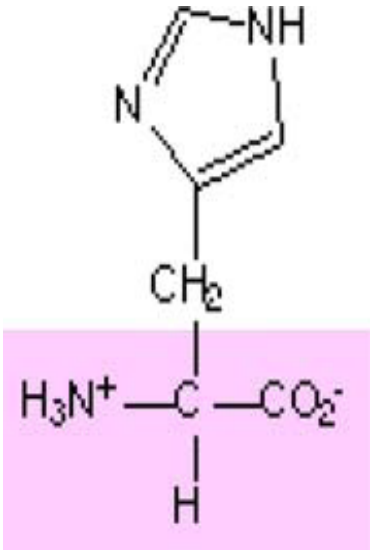
Codons	Anticodons
UAU	AUA
UAC	GUA



Tyr-specific RNA aptamers:

ggcAGucaacucgugcgaucgugaaaAcGGGGcaAGAuGGccuuAcaGCG
GUCAAUACGGGGGuCAGAUAGGGGAGGCCUcCUggu

C. Mannironi et al 2000



Histidine:

codons	anticodons
CAU	AUG
CAC	GUG

His-specific RNA aptamers:

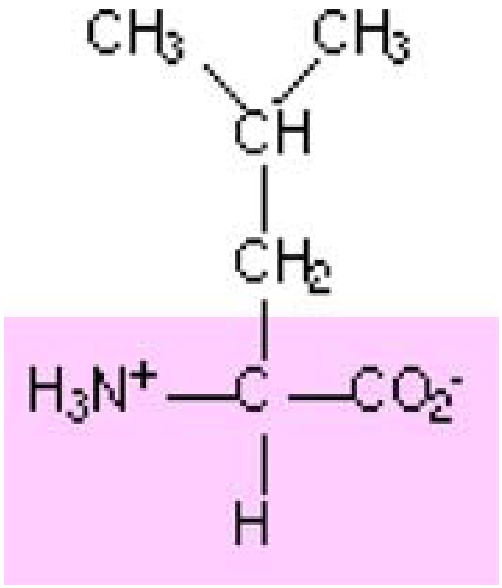
 # His 17:

cagcAAGCGGGGaaaAUGUuGGgAACAGcugcggaaggaaaucauggagg…

 # His729:

uacaAAGUGGAuGAGUuAGgAACAGguauuuaugcauggguggaguucgg…

Majerfeld et al., 2005



Leucine

Codons	Anticodons
UUA	UAA
UUG	CAA
CUU	AAG
CUC	GAG
CUA	UAG
CUG	CAG

Leu-specific RNA aptamers:

 # Leu 112:

ucucucAacccc**UAG**cgUAgUUUUGAcUGcGAGAGGCAAAcg
ccacgg**UAG**AACCGAagGG**UAG**gagggauua

Majerfeld et al., 2005

Summary

- For amino acids encoded by dinucleotide-palindrome-containing triplets, their binding sites in RNA aptamers "prefer" the codon(1-2)/anticodon(2-3) motifs over codon(2-3)/anticodon(1-2) counterparts in spite of their seemingly perfect symmetry.
- Anticodons "drive" aa-specificity, codons being trivial hitch-hikers.

These striking preferences mean that the 3rd nt is more important than 1st nt in anticodons (complementarily, the reverse being the case of codons) which is ***precisely as in the real genetic code*.** However, since selection of aa-specific RNA aptamers apparently had nothing to do with translation, it would be correct to say that in the interface between interacting aas and cognate triplets, the 2-3 nucleotides contribute more to the specificity of interaction thus determining their future usage as an **anti**codon.

- Primordial r-aaRSs could simply have used the preexisting aa-triplets affinities in the way that minimized errors of aminoacylation. That is, the original aa-triplet preferences within the aa-binding sites of RNA catalysts determined the primal pre-translational genetic code with more important 2nd and 3rd nts, whereas 1st nt was much less specific.
- Later, when the code was expanding in co-evolution with the translation apparatus, the importance of 2-3 nts of coding triplets passed on 1-2 nts of their complements thus distinguishing anticodons from codons. The fact that codon's 3rd nt is more degenerated than anticodon's 1st nt serves as an indirect evidence in favor of this order of events.
- Does this translation-independent preference of (1-2)/(2-3) over (2-3)/(1-2) triplets suggest some fundamental left-right, **chirality-like**, asymmetry? and if it does, could this asymmetry determine the salient features of coding and translation?

Conventional wisdom stipulates that code shaping:

- recapitulates amino acids biosyntheses,
- was directed by minimization of translation errors, etc.
- reflects its co-evolution with p-aaRSs

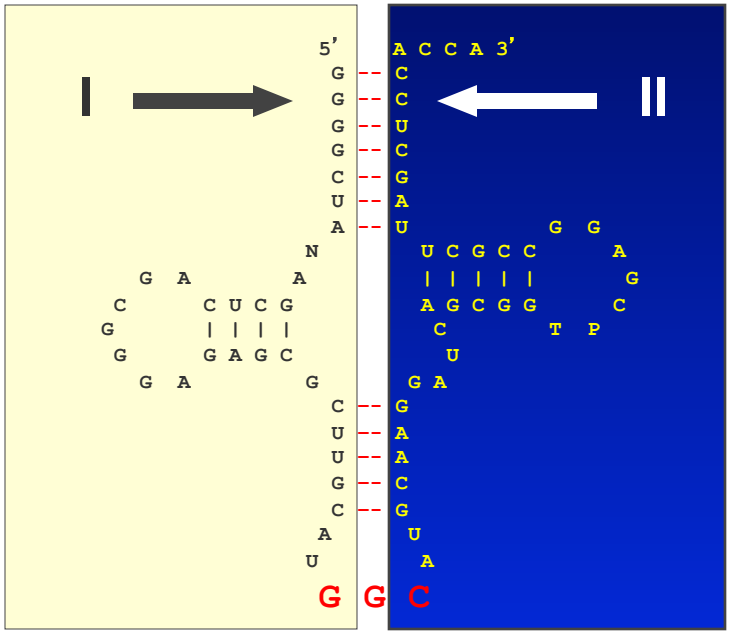
While not completely denying the role of translation-motivated co-evolution in final code shaping, we bring attention to the reverse flow of causality in the earliest code shaping stage.

Not only the translation machinery itself, but its coding tool kit, the code adapting (tRNA) and code implicating (r-aaRSs) molecules, have been evolving to ‘‘fit’’ the probably already existing code (rather than the code co-evolving with tRNAs and aaRSs to fit translation).

Translation without code does not make sense, code without (before) translation does.

Part 2

Imprints of the code in tRNAs and aaRSs



10 : 10

There are two modes of tRNA recognition by aaRS:

from opposite (minor and major groove) sides of the acceptor with attaching the cognate aa to 3' OH and 2' OH hydroxyls of A76, respectively.

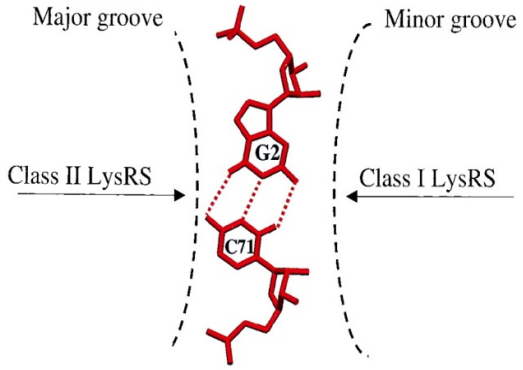
Why?

Palindromes CG, GC, UA, AU provide an answer

5’	U		C		A		G		3’
	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
U	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	*Stop*	UGA	*Stop*	A
	UUG	Leu	UCG	Ser	UAG	*Stop*	UGG	Trp	G
	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
C	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
A	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
G	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

aaRS I

aaRS II



Ibba et al., 1997

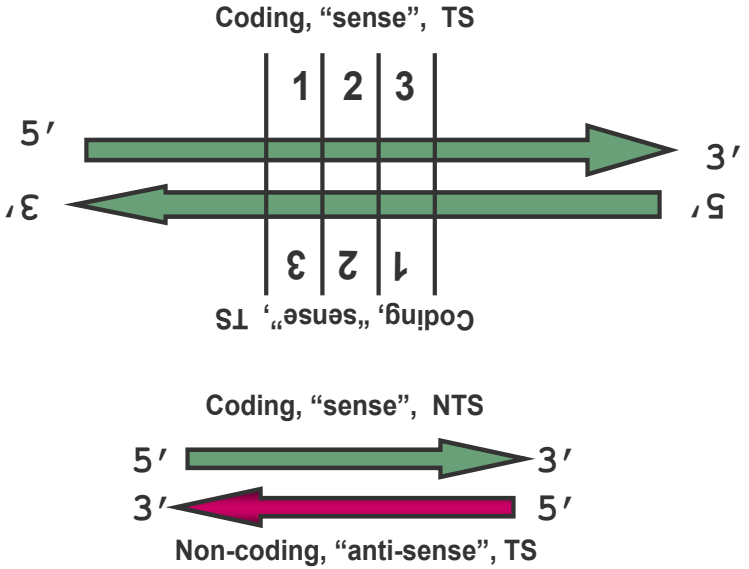
The double assignment of LysRS hints that either of the two enzyme classes is probably versatile enough to be able to aminoacylate tRNAs ***in all 20 cases***. **Then why are aaRSs of two types?**

1	2				3
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
U	Phe	Ser	Tyr	Cys	C
U	Leu	Ser	Stop	Stop	A
U	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
C	Leu	Pro	His	Arg	C
C	Leu	Pro	Gln	Arg	A
C	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
A	Ile	Thr	Asn	Ser	C
A	Ile	Thr	Lys	Gly/Ser	A
A	Met	Thr	Lys	Gly/Ser	G
G	Val	Ala	Asp	Gly	U
G	Val	Ala	Asp	Gly	C
G	Val	Ala	Glu	Gly	A
G	Val	Ala	Glu	Gly	G

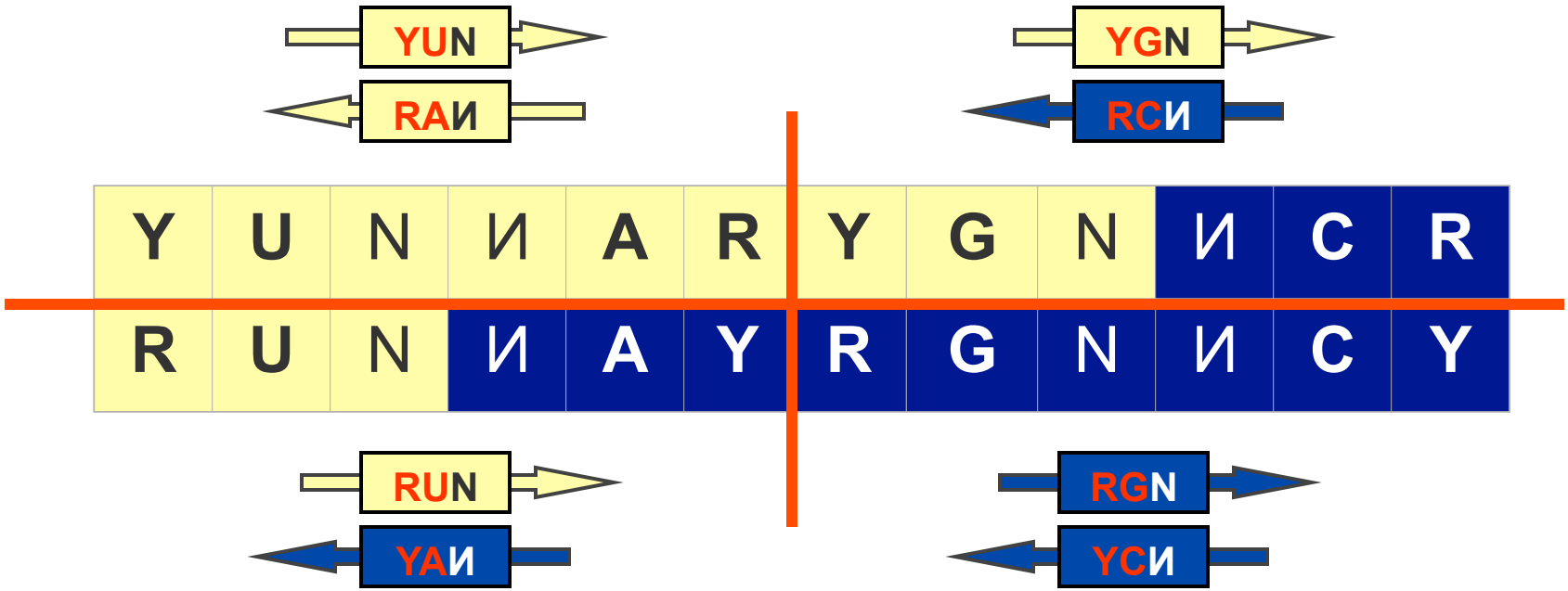
Minor groove side

Major groove side

DOUBLE STRAND CODING



1	2 U	3	1	2 A	3	1	2 G	3	1	2 C	3
U	Phe	U	A	*Lys*	A	U	Cys	U	A	Thr	A
U	Phe	C	G	Glu	A	U	Cys	C	G	Ala	A
U	Leu	A	U	Stop	A	U	Stop	A	U	Ser	A
U	Leu	G	C	Gln	A	U	Trp	G	C	Pro	A
C	Leu	U	A	*Lys*	G	C	Arg	U	A	Thr	G
C	Leu	C	G	Glu	G	C	Arg	C	G	Ala	G
C	Leu	A	U	Stop	G	C	Arg	A	U	Ser	G
C	Leu	G	C	Gln	G	C	Arg	G	C	Pro	G
A	Ile	U	A	Asn	U	A	Ser	U	A	Thr	U
A	Ile	C	G	Asp	U	A	Ser	C	G	Ala	U
A	Ile	A	U	Tyr	U	A	*Gly/Ser*	A	U	Ser	U
A	Met	G	C	His	U	A	*Gly/Ser*	G	C	Pro	U
G	Val	U	A	Asn	C	G	Gly	U	A	Thr	C
G	Val	C	G	Asp	C	G	Gly	C	G	Ala	C
G	Val	A	U	Tyr	C	G	Gly	A	U	Ser	C
G	Val	G	C	His	C	G	Gly	G	C	Pro	C

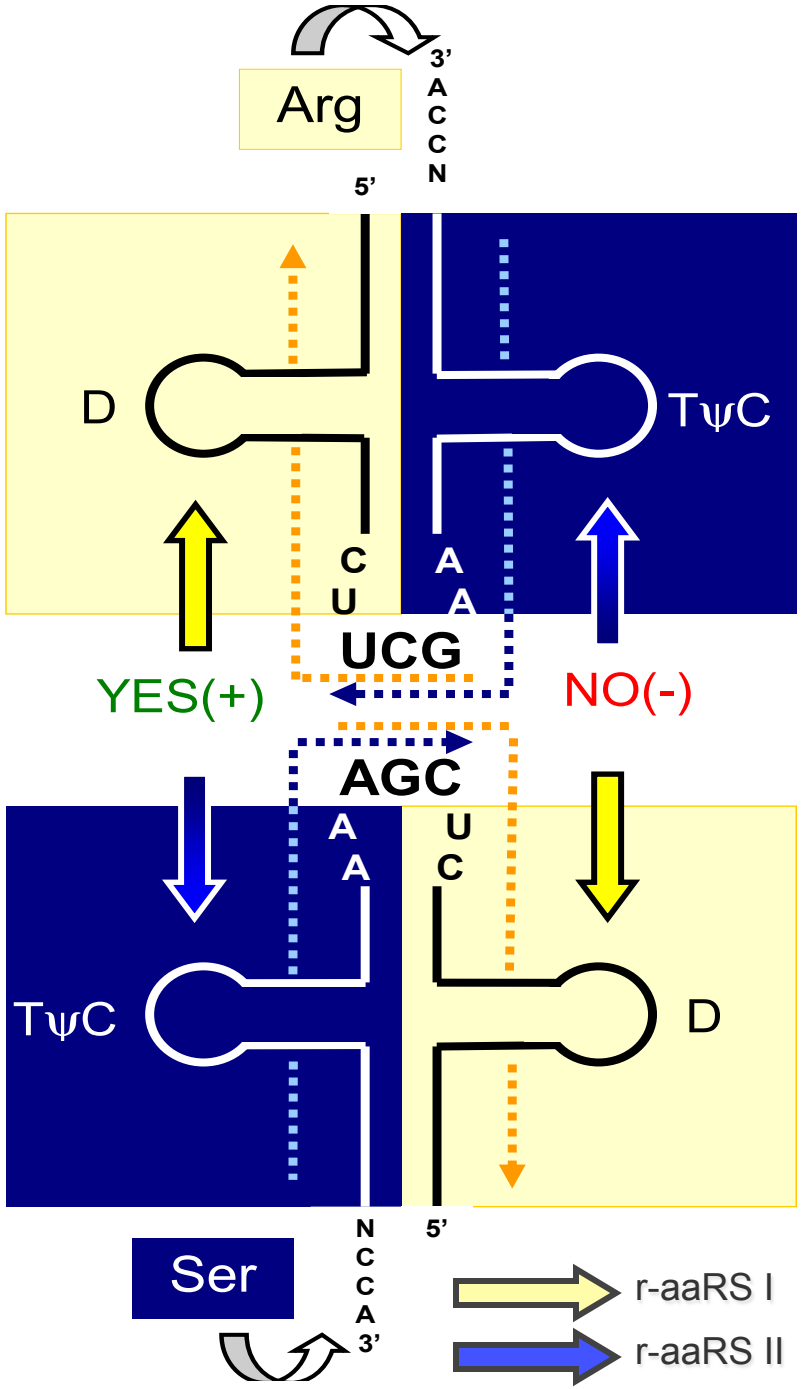


Sub-code for two modes of tRNA recognition:

- If two complementary codons have two purines versus two pyrimidines (**RR** vs. **YY**) at the neighboring positions (1st and 2nd vs. 2nd and 3rd), the corresponding amino acids belong to the same class --- class I for N**AR** vs. **YU***И* codon pairs, and class II for **RG**N vs. *И***CY** codon pairs.
- If, in a pair of complementary codons, these two adjacent positions are occupied by a purine and a pyrimidine (**YR** vs. **YR** or **RY** vs. **RY**), the corresponding amino acids belong to the different classes --- class I **YG**N vs. class II *И***CR** (upper right), mirrored by class I **RU**N vs. class II *И***AY** (lower left).

Each of CG, GC, UA and AU dinucleotides is a perfect palindrome indistinguishable from its complement.

Rodin & Rodin, 2006, 2008



5'-cUXYZAN-3'

r-ArgRS

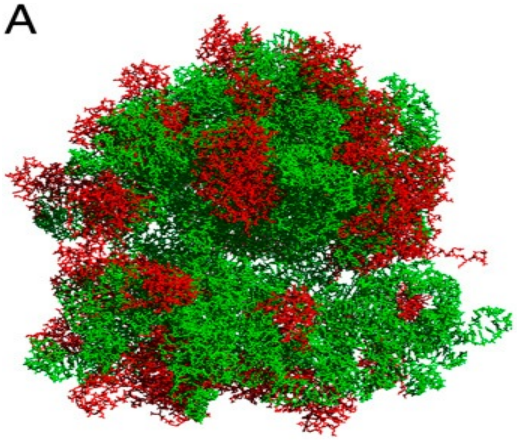
5'-CUUCGAA-3'

5'-CUCGAAA-3'

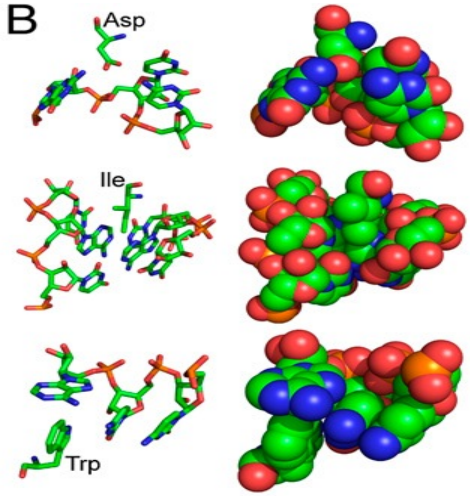
r-SerRS

The yin/yang-like sub-code for two modes of tRNA aminoacylation minimizes the risk of confusion of tRNAs with complementary anticodons

Since r-aaRSs recognized the complementary halves of proto-tRNAs, their tRNA-binding sites are supposed to have been complementary to each other as well. *(Rodin & Rodin, 2006, 2008)*
To simplify matters, one would assume for proto-tRNAs with complementary anticodons a kind of aa-specific cross-self-aminoacylation activity (located for instance within introns, right after the nearly invariant 37th nucleotide – adenine) *(Rodin and Ohno, 1997).*



The structure of *Thermus thermophilus* ribosome with proteins highlighted in red and rRNAs in green.



Examples of amino acid contacting its anticodon rRNA shown in the stick (*Left*) and spherical (*Right*) representation.

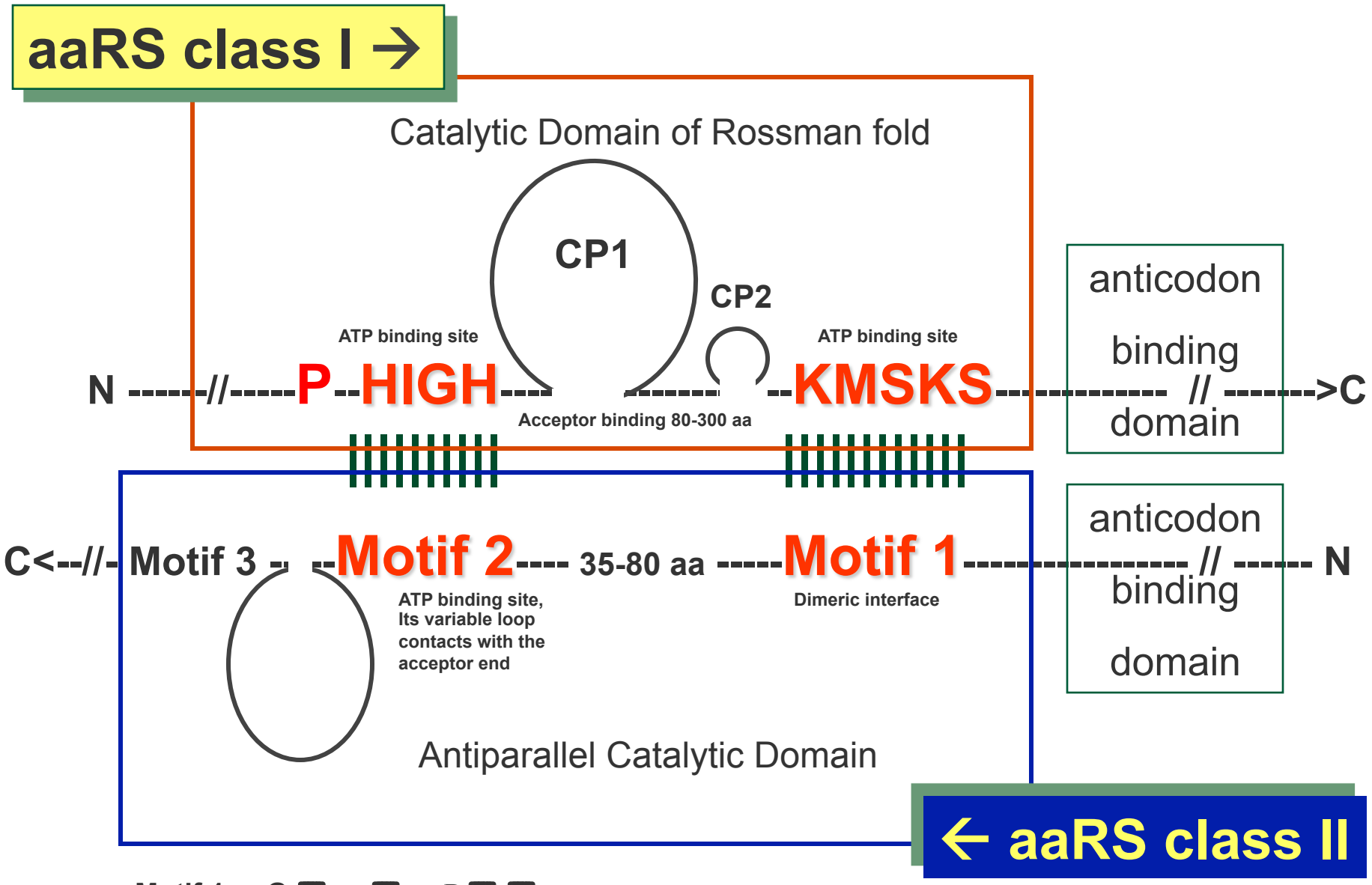
Imprints of the code in the ribosome?

"…We show here that anticodons are selectively enriched near their respective amino acids in the ribosome, and that such enrichment is significantly correlated with the canonical code over random codes. The ribosome thus serves as a molecular fossil, preserving biological evidence that anticodon – amino acid interactions shaped the evolution of the genetic code."

Johnson D. & Wang L, 2010. PNAS USA 107: 8298–8303

- the complementarity of the two putative r-aaRSs, and
- **analyses of aa-binding sites of "selexed" RNA aptamers (suggesting a certain stereo-chemical affinity between aa and anticodons) bring us to the hypothesis that**

The very first two precursors of minimalist p-aaRSs (that replaced complementary r-aaRSs and interacted with complementary "anticodons") have also originated from the complementary strands of one and the same ancestral gene, which is exactly what we observe.

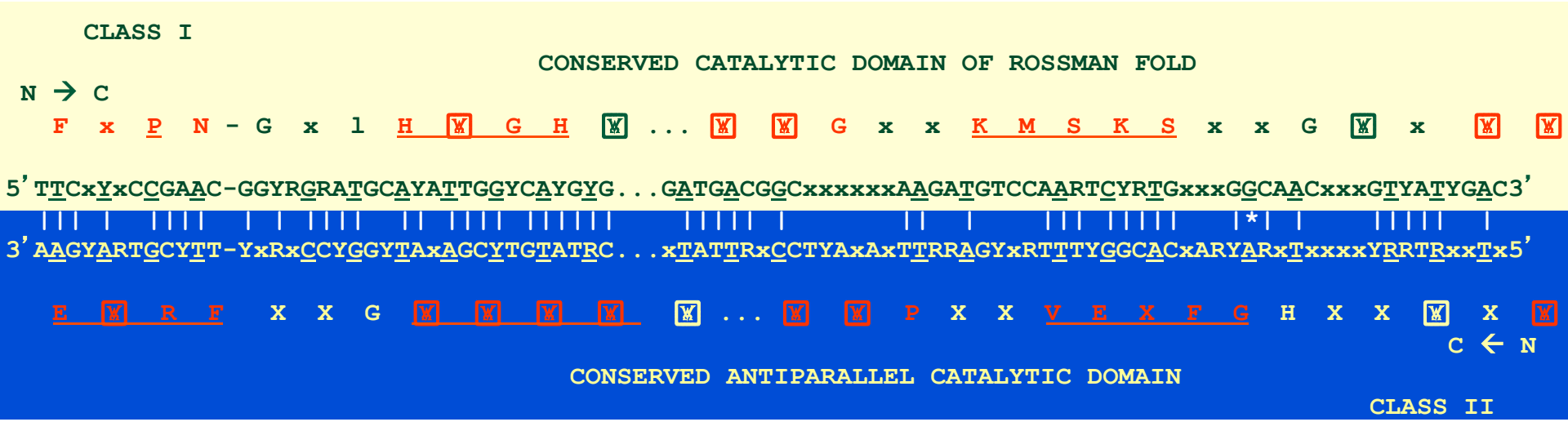


Motif 1: + G ▯ xx ▯ xx P ▯ ▯

Motif 2: + ▯ ▯ +/- ▯ xxx FR x E/D… … + ▯ xx – F xxx - ▯ x ▯ ▯

Motif 3: G ▯ G ▯ G ▯ E R ▯ ▯ ▯ ▯ ▯

Rodin & Ohno, 1995



Rodin & Ohno, 1995

Updated in Rodin, Rodin & Carter 2009

Preference of codon1-2/anticodon2-3 motifs and primacy of anticodons in aa-binding sites of "selexed" RNAs, plus

- The concerted dual complementarity: tRNAs with complementary anticodon have complementary 2nd bases in their acceptor stems *(Rodin, Rodin & Ohno, 1996; updated in Rodin, Szathmáry, Rodin, 2009)*
- The sub-code for two modes of tRNA recognition by aaRSs – from minor and major groove sides of the acceptor stem *(Rodin & Rodin, 2006, 2008)*
- The SAS origin of these two aaRSs inherited from aminoacylating ribozymes *(Rodin & Ohno, 1995)*
- The "Fibonacci" model of tRNA growth from anticodon triplet and DCCA tetraplet to a 76nt-long cloverleaf *(Rodin, Szathmáry, Rodin, 2011).*

… make the following hypotheses

- stereo-chemical affinity between amino acids and cognate triplets *(Woese, 1967; Orgel, 1968; Yarus, 1988-2011),*
- tRNA-like genomic 3' tags suggesting that tRNAs originated in replication *(Weiner & Maizels, 1987, 1994),*
- ancient ribozymes-mediated operational code of proto-tRNA aminoacylation *(Hou & Schimmel, 1988; Schimmel et al., 1993-2011; De Duve, 1988) …*

… not mutually contradictory, but rather co-existing in harmony.

Acknowledgements

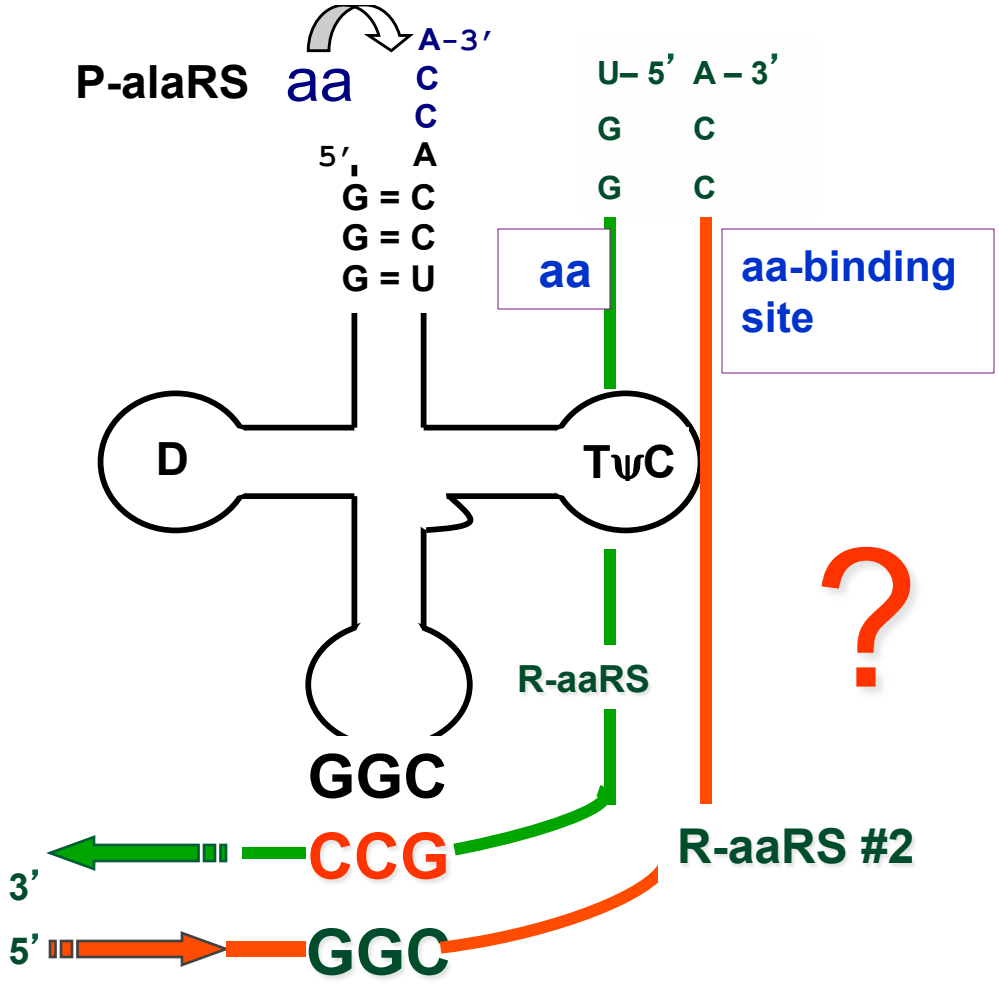
- **Andrei Rodin** (Human Genetics Center, University of Texas, Houston, USA)
- **Eörs Szathmáry** (Collegium Budapest, Hungary)
- **Paul Schimmel** (Scripps Institute, La Jolla, CA, USA)
- **Charles Carter** (University of North Carolina, Chapel Hill, USA)
- **Massimo Di Giulio** (Institute of Genetics and Biophysics, Naples, Italy)
- **Germinal Cocho** (Institute of Theoretical Physics, UNAM, Mexico-City, Mexico)
- **Apoorva Patel** (Centre for High Energy Physics, Bangalore, India)
- **Koji Tamura** (Tokyo University of Sciences, Japan)
- **Mark Safro** (Weizmann Institute of Sciences, Rehovot, Israel)
- **Lluis Ribas de Pouplana** (Biomed.Res.Inst, Barcelona)

- **Eugene Roberts** (City of Hope, Duarte, CA, USA)
- **John Rossi** (City of Hope, Duarte, CA, USA)
- **Arthur Riggs** (City of Hope, Duarte, CA, USA)

R–aaRSs as a solution ?

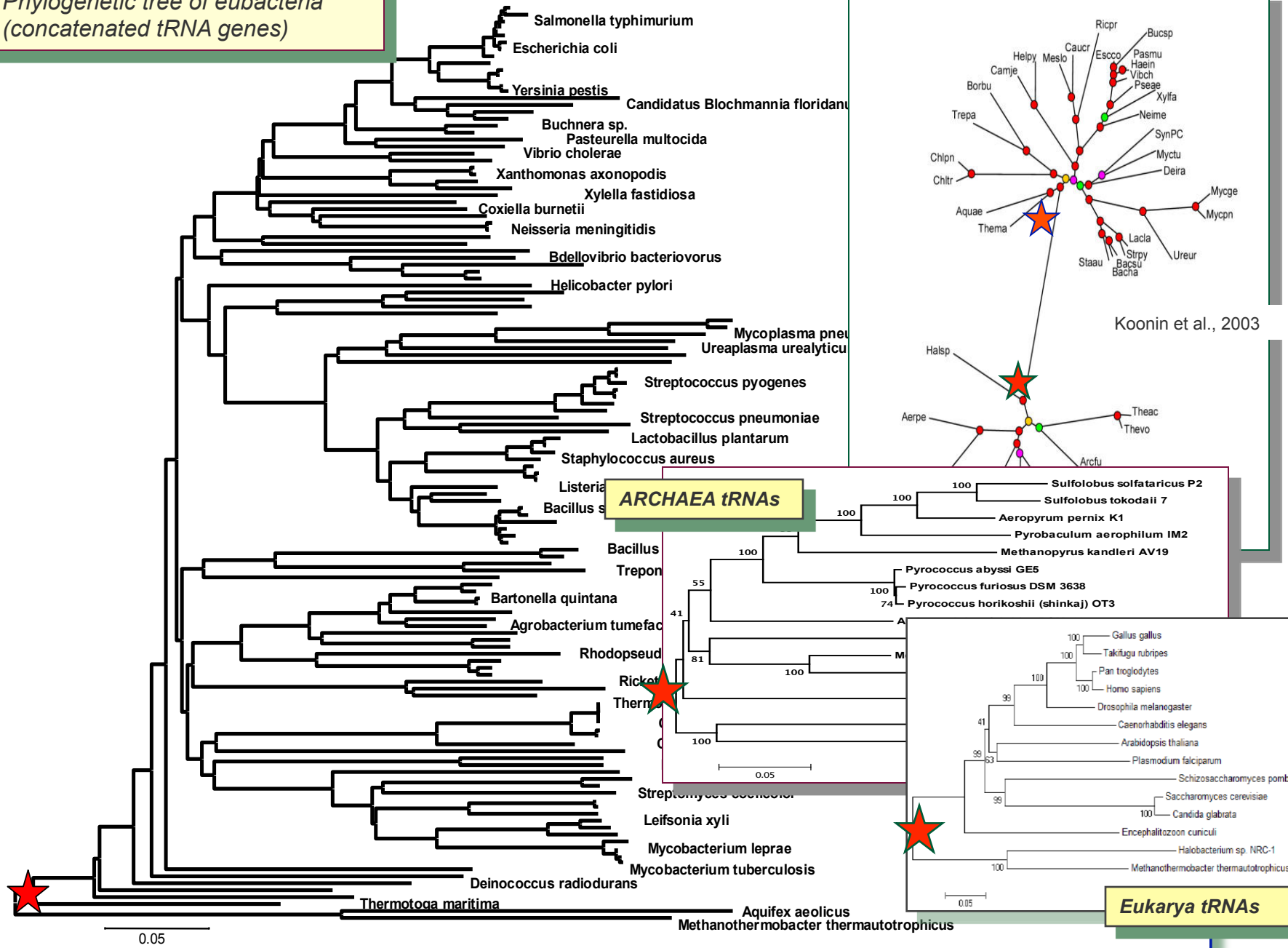
R–aaRSs had obvious advantage over their protein successors (P–aaRS) --- an ability to precisely recognize the remotely located anticodon through complementary pairing (i.e. simply by the corresponding codon-like triplet).

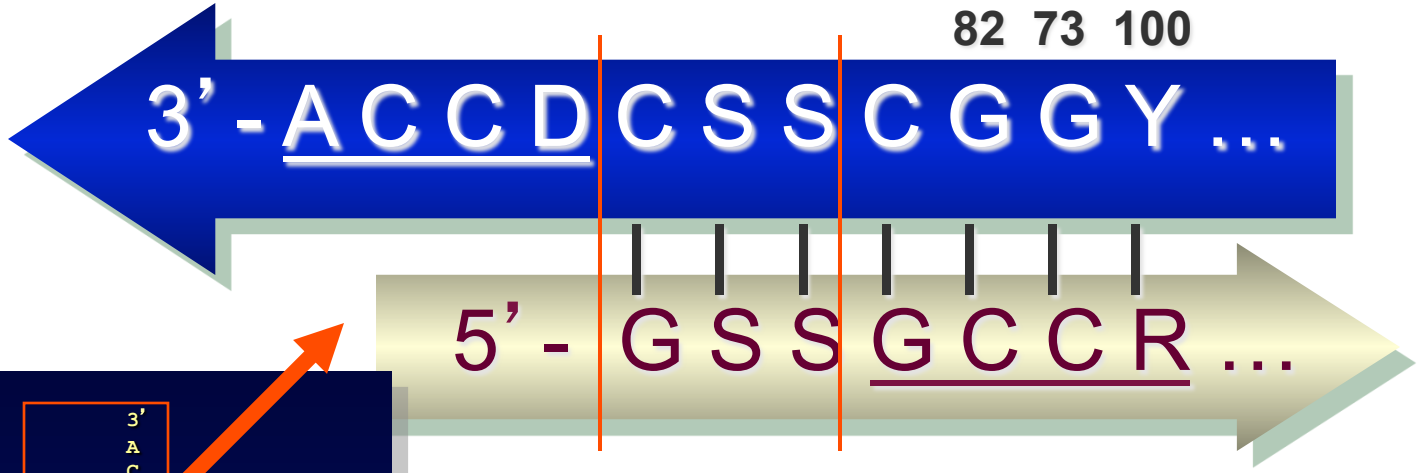
However…in order to attach the appropriate amino acid to "its" tRNA, the R–aaRS also must have had the specific aa-binding site, but positioned as close as possible to the tRNA 3' end in the R-aaRS – tRNA complex.

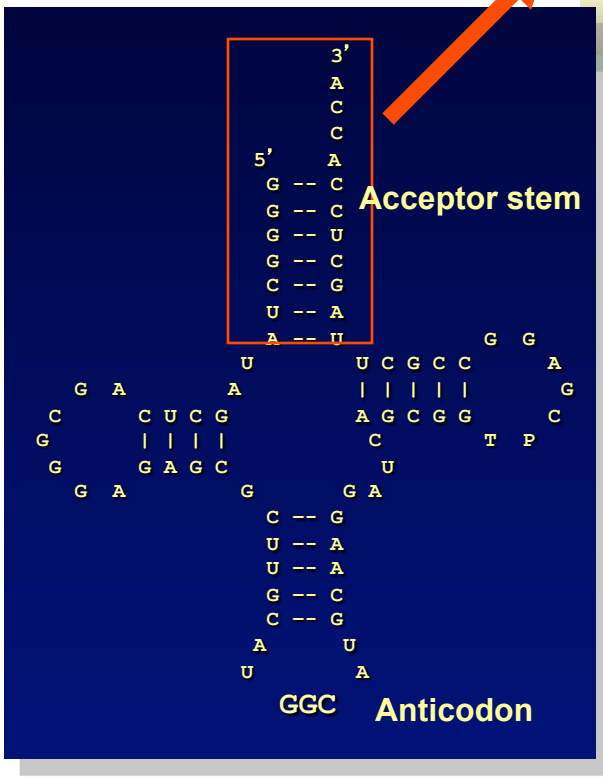


The situation is symmetric: we re-target the problem of the remote location of the anticodon (codon)-specific aa-binding site from tRNA to its hypothetical (topologically more complex and bulkier) aminoacylating ribozyme, R-aaRS, which in order to aminoacylate their cognate tRNAs, would require catalysts of their own, i.e., "meta-r-aaRSs", which, in turn, would inherit the same problem and require catalysts of their own, etc...*ad infinitum*.

Phylogenetic tree of eubacteria (concatenated tRNA genes)







R = G or A
Y = C or U
S = G or C
D, base-determinator (mostly A, G)

(Rodin, Rodin & Ohno, 1996; 1268 tRNA genes
Rodin, Szathmary & Rodin, 2009; > 8000 tRNA genes)

3’–ACCACCCCSGC…
| | | | | | |
5’–GGGGSCG…

(Eigen et al., 1989; ~1000 tRNA genes)

Anticodon GSS(ssC):3

DCCA(UGGd):4

3 + 4 = 7

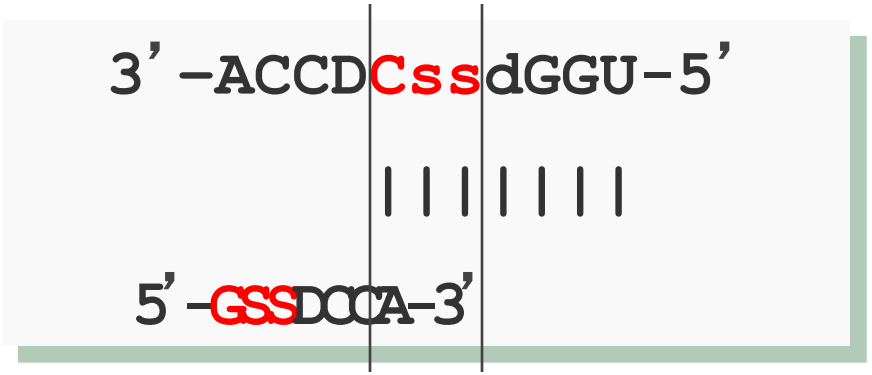
4 + 7 = 11

7 + 11 = 18

11 + 18 = 29

29 + 18 = 47

29 + 47 = 76



7/4, 11/7, 18/11, 29/18, 47/29, 76/47:

$(1 + \sqrt{5})/2 = 1.618$

Building units
Coding triplets: 5'-**GCC**-3' and 5'-**GGC**-3'
Flanking tetraplets: 5'-**DCCA**-3' and 5'-**UGGd**-3'

Elongation by self-priming and self-templating

1st step: 4 + 3 = 7

3′-ACCD-5′+ 3′-CCG-5′ -> 3′-ACCDCCG-5′

2nd step: 7 + 4 = 11

3′-ACCDCCG-5′+ 3′-dGGU-5′ → 3′-ACCDCCGdGGU-5′

3′-ACCDCCGdGGU-5′
↕
5′-UGGdGGCDCCA-3′

3rd step: 11 + 7 = 18

5′-GGCDCCA-3′
3′ACCDCCGdGGU-5′ -> 3′-ACCDCCGdGGU…ACCDCGG-5′ ->

3′-ACCDCCGdGGU…
|||||||
5′-GGCDCCA…

4th step: 18 + 11 = 29

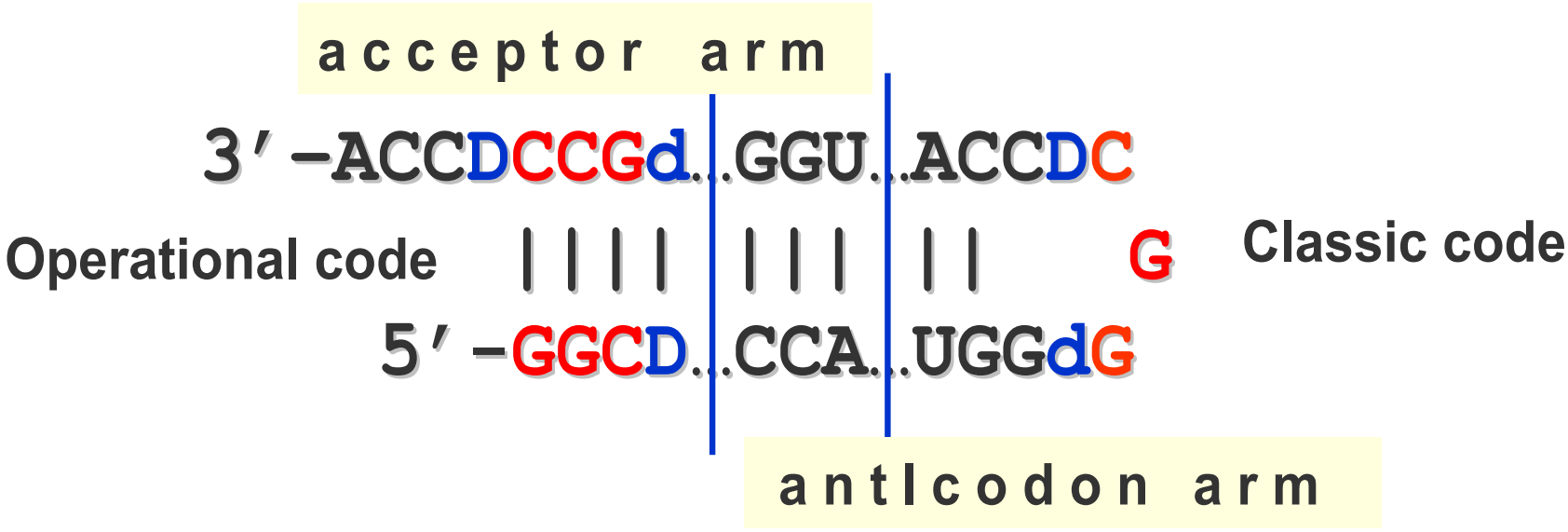
5′-UGGdGGCDCCA-3′
3′-ACCDCCGdGGU-5′ …ACCDCGG-5′-> 3′-ACCDCCGdGGU…ACCDCGGdGGU…ACCDCGG-5′:

3′-ACCDCCGdGGU
|||||||
5′-GGCDCCA

+ 3′-ACCDCGGdGGU-5′ ->

C D
3′-A C C D C C G d G G U A C C
| | | | | | | | | | | G
5′ - G G C D C C A U G G
G d

RNA world → RNA + proteins



5th step: 29 + 18 = 29 + (11 + 7) = 47

```
3′-ACCDCCGdGGU…ACCDC          3′-ACCDCCGdGGU-5′
       |||||||  ||    G   +      |||||||          :
    5′-GGCDCCA…UGGdG          5′-UGGdGGC-3′
```

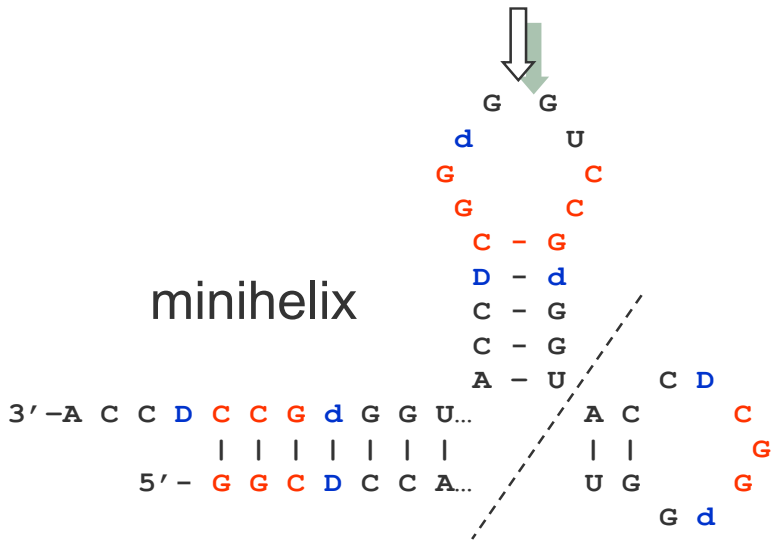
```
5′-UGGdGGCDCCA-3′
3′-ACCDCCGdGGU……ACCDCGGdGGU…ACCDCGG-5′
```

↓

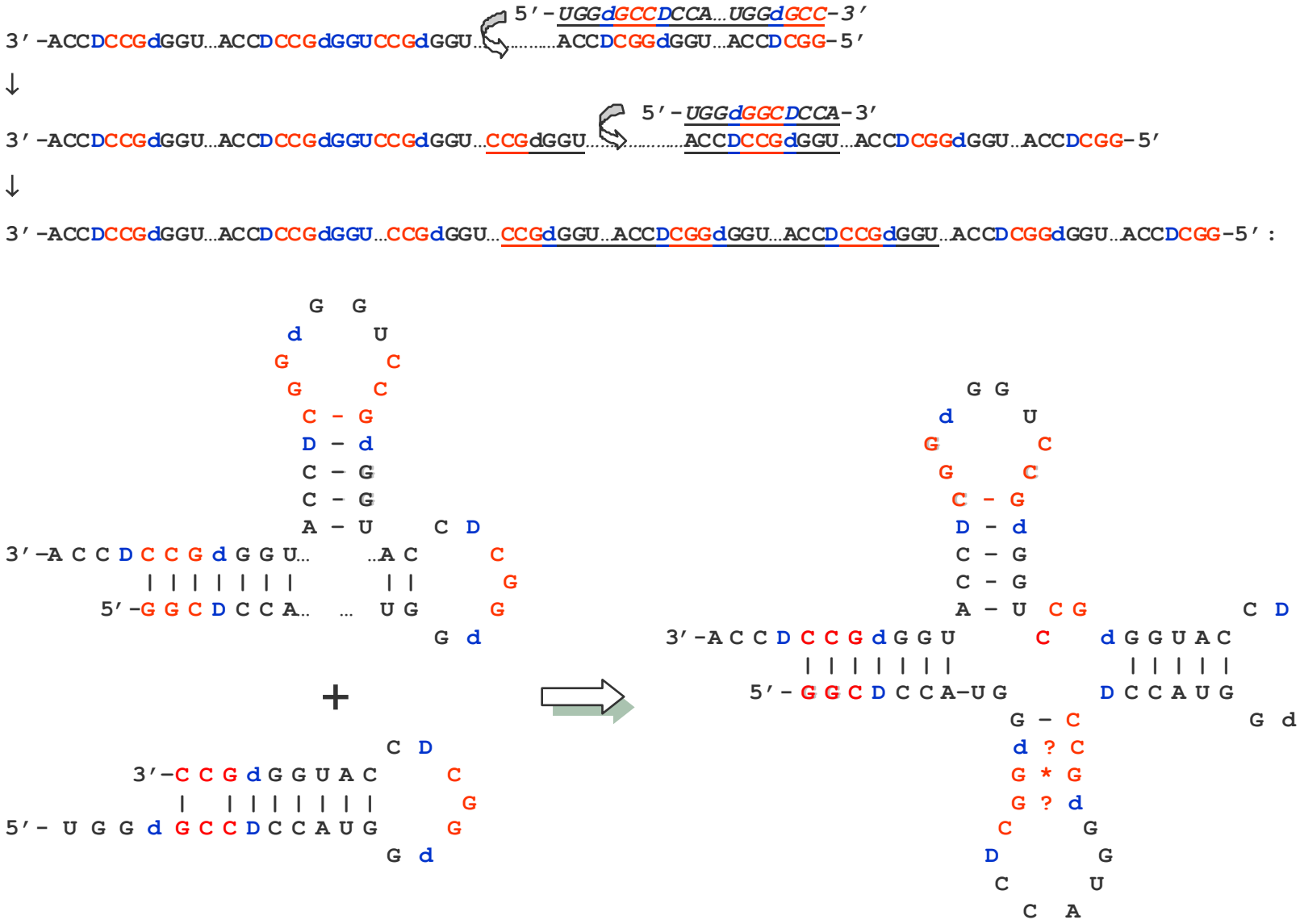
```
          5′UGGdGCC-3′
3′-ACCDCCGdGGU…ACCDCGGdGGU……ACCDCGGdGGU…ACCDCGG-5′
```

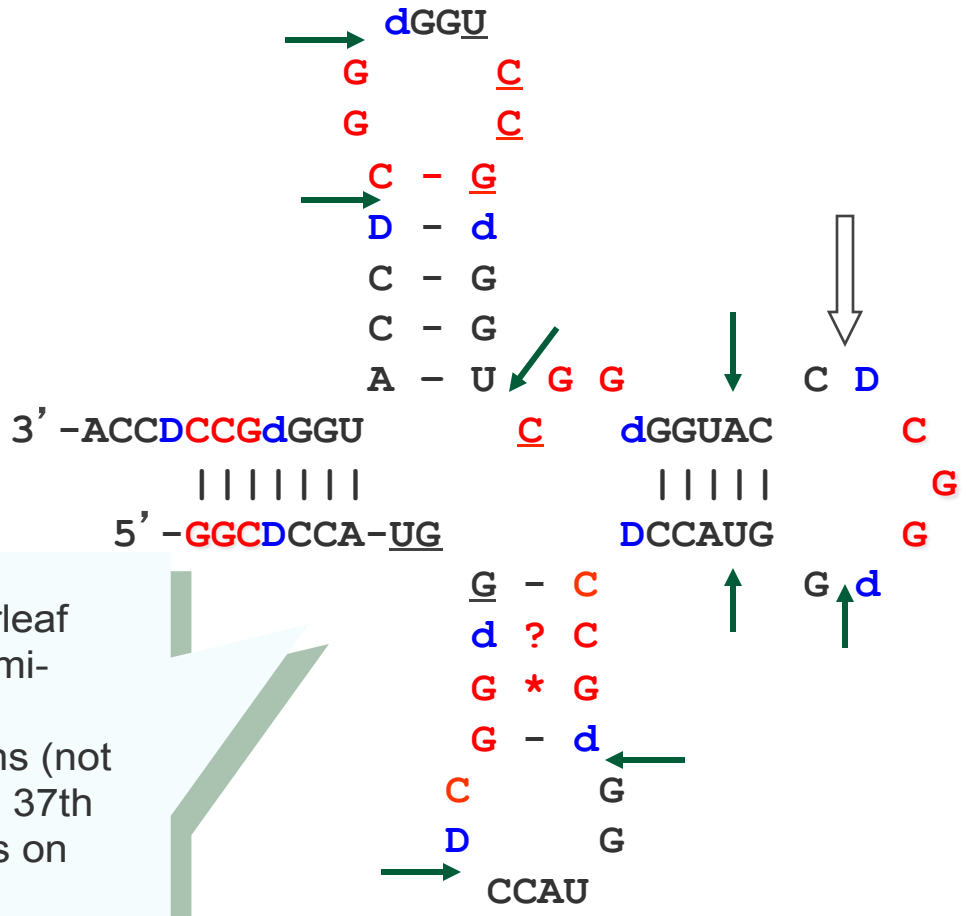
↓

```
3′-ACCDCCGdGGU…ACCDCCGdGGUCCGdGGU…ACCDCGGdGGU…ACCDCGG-5′
```



6th step: 47 + 29 = 47 + (18 + 11) = 76





This reconstructed tRNA cloverleaf reproduces all invariant and semi-invariant nucleotides, with characteristic locations of introns (not only the major site between the 37th and 38th nucleotides), the splits on minigenes in archaeal parasite *Nanoarchaeum equitance* (*Söll et al., 2005*), and the positions of processing in permuted tRNA genes from red algae *Cyanidioschyzon merolae* (*Soma et al., 2007).*

Building units:

Coding triplets: 5'-GGC-3' and 5'-GCC-3'

Flanking tetraplets: 5'-DCCA-3' and 5'-UGGd-3'

proto-tRNA?

A: aminoacylation

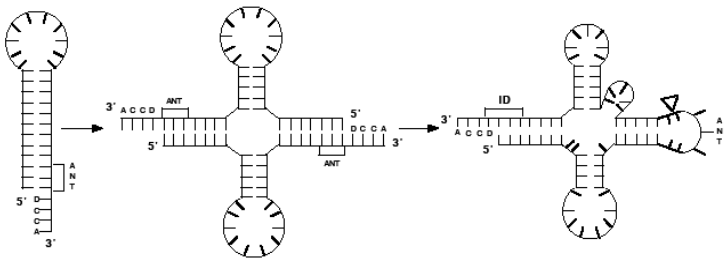
```
3'-ACCDCCGdGGU-5'
       |||||||
    5'-GGCDCCA-3'
```

B: translation?

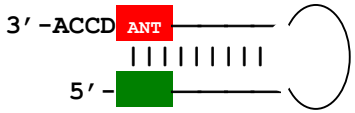
```
3'-ACCDCCGACCDC
          |||   G
       5'-UGGdG
```

C: replication?

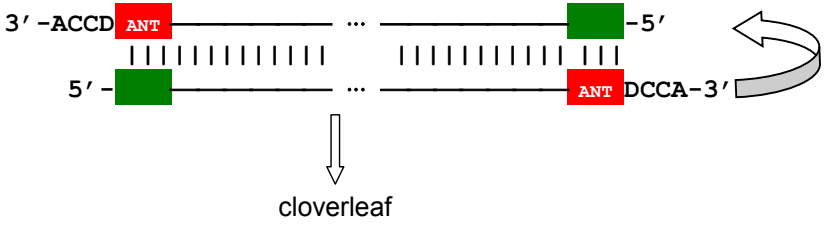
```
       3'-ACCDC
          |||   G
5'-GCCDCCAUGGdG
```



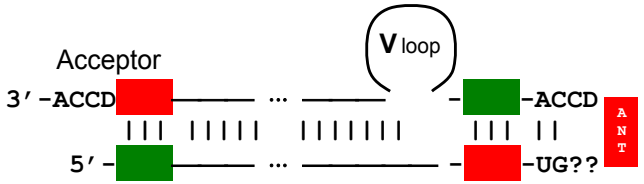
Original hairpin



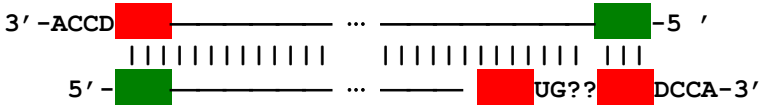
Dimerization



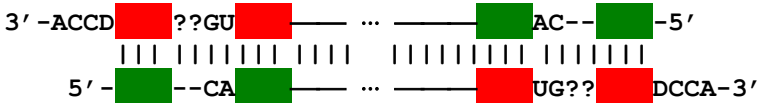
…implies:



…implies that in the original hairpin:



…implies:



ARGININE

Anticodon GCG

Codon CGG

Overrepresented

Arg

5′---GCGG-->3′

3′<--CGCC---5′

Ala?

Overrepresented?

Vs.

Anticodon GCG

Codon CGC

Absent

Arg

5′---GCGC-->3′

3′<--CGCG---5′

Ala?

Absent?

ISOLEUCINE

Anticodon UAU

Codon AUU

Overrepresented

Ile

5′---UAUU-->3′

3′<--AUAA---5′

Tyr

Overrepresented

Vs.

Anticodon UAU

Codon AUA

Absent

Ile

5′---UAUA-->3′

3′<--AUAU---5′

Tyr

Absent

RNA world (ribozymes) → RNP world (enzymes) (origins of code and translation)

One cannot simply refer to the truism that proteins are more efficient and versatile catalysts than RNAs since any such advantage appears in the end of ribozymes-mediated multi-step coding and translation processes.

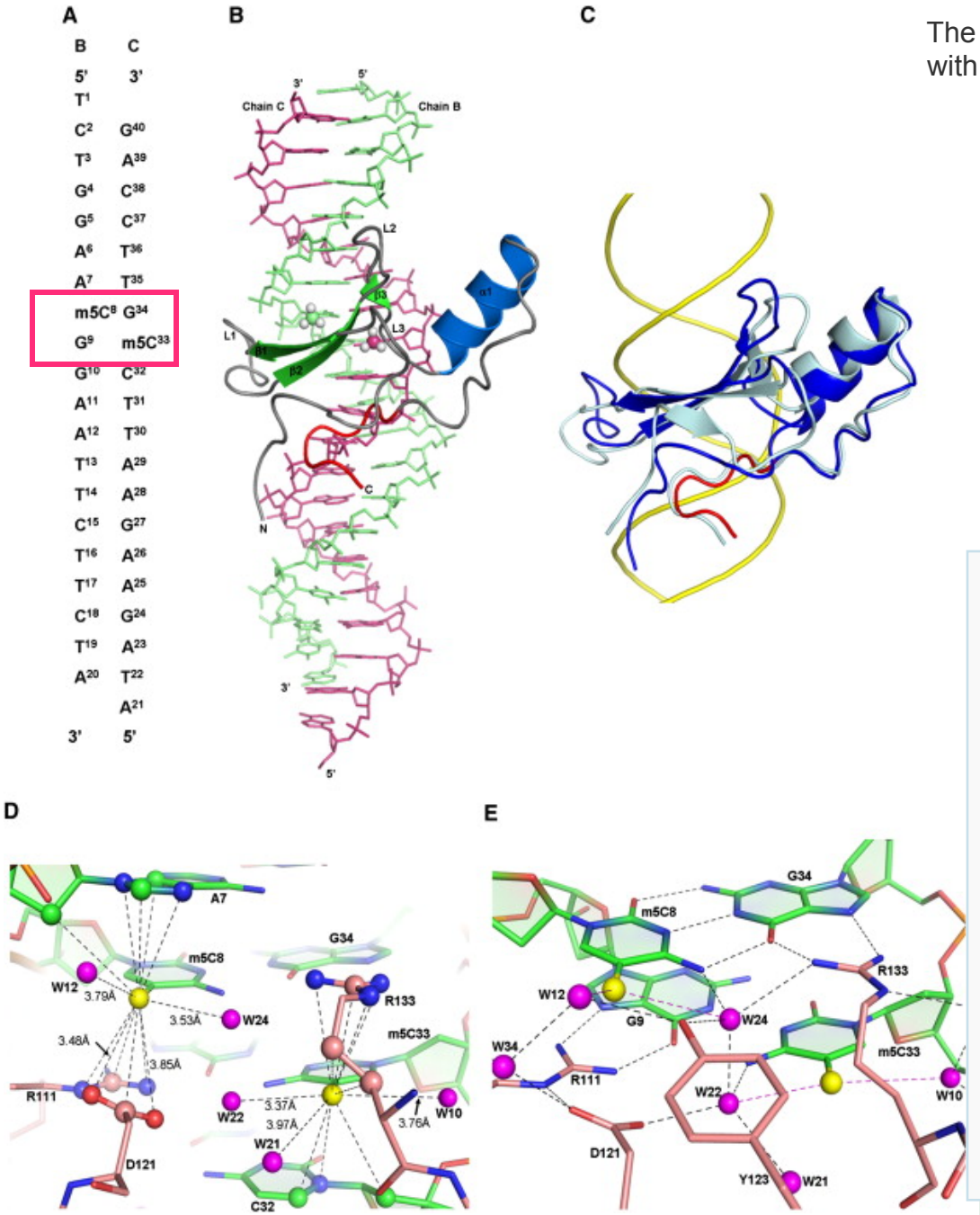
However, selection works like a first-aid ambulance, not in foresight of future demands.

As in any case of step-by-step evolution towards a more complex system, we have to propose Darwinian explanation to each step.

First of all, it seems logical (1) to separate the two origins (origin of the code and origin of translation) and (2) to stipulate that **the code emerged before translation** – in response to the demands of the RNA life.

Basic premise: *Translation without code does not make sense, code without (before) translation does.*

The X-Ray Structure of MeCP2-MB complexed with *BDNF* Promoter DNA at 2.5 Å



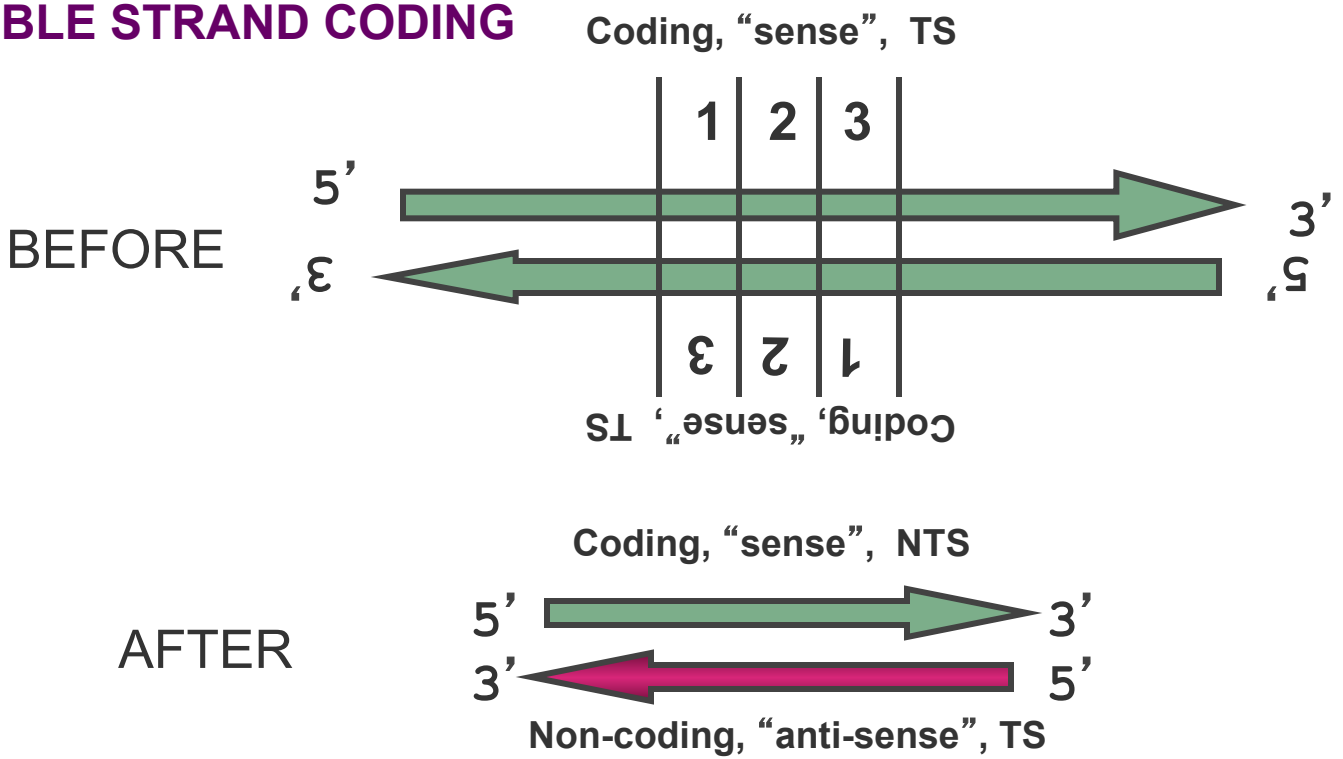
In the methyl-CpG-binding domain (MBD) of transcriptional repressors such as MeCP2-MBD co-crystallized with its specific promoters, the only residues that do form direct hydrogen bonds with guanines of the palindromic m5CpG pair are arginines R111 and R133 *(Ho et al., 2008)*, and this interaction seems to be essential for diverse examples of DNA-protein recognition *(Luscombe et al., 2001)*.

The "vicious circle" controversy:

- The genetic code originated in the RNA world.
- Initially, proto-tRNAs were aminoacylated by R-aaRSs.
- P-aaRSs appeared later and (strikingly) in two versions.
- Modern I and II P-aaRSs (and most likely their ancient precursors) recognize(d) the acceptor stem rather than the anticodon, thus directly contributing to the operational RNA code, and (only) indirectly – to the genetic code proper…
- --- and yet, the genetic code itself displays signs of a nonrandom distribution of class I and II P-aaRSs among amino acids (*Ribas De Pouplana & Schimmel, 2001, Rodin & Rodin, 2006, Delarue 2007, Pham et al., 2007*)…→ the circle is complete.

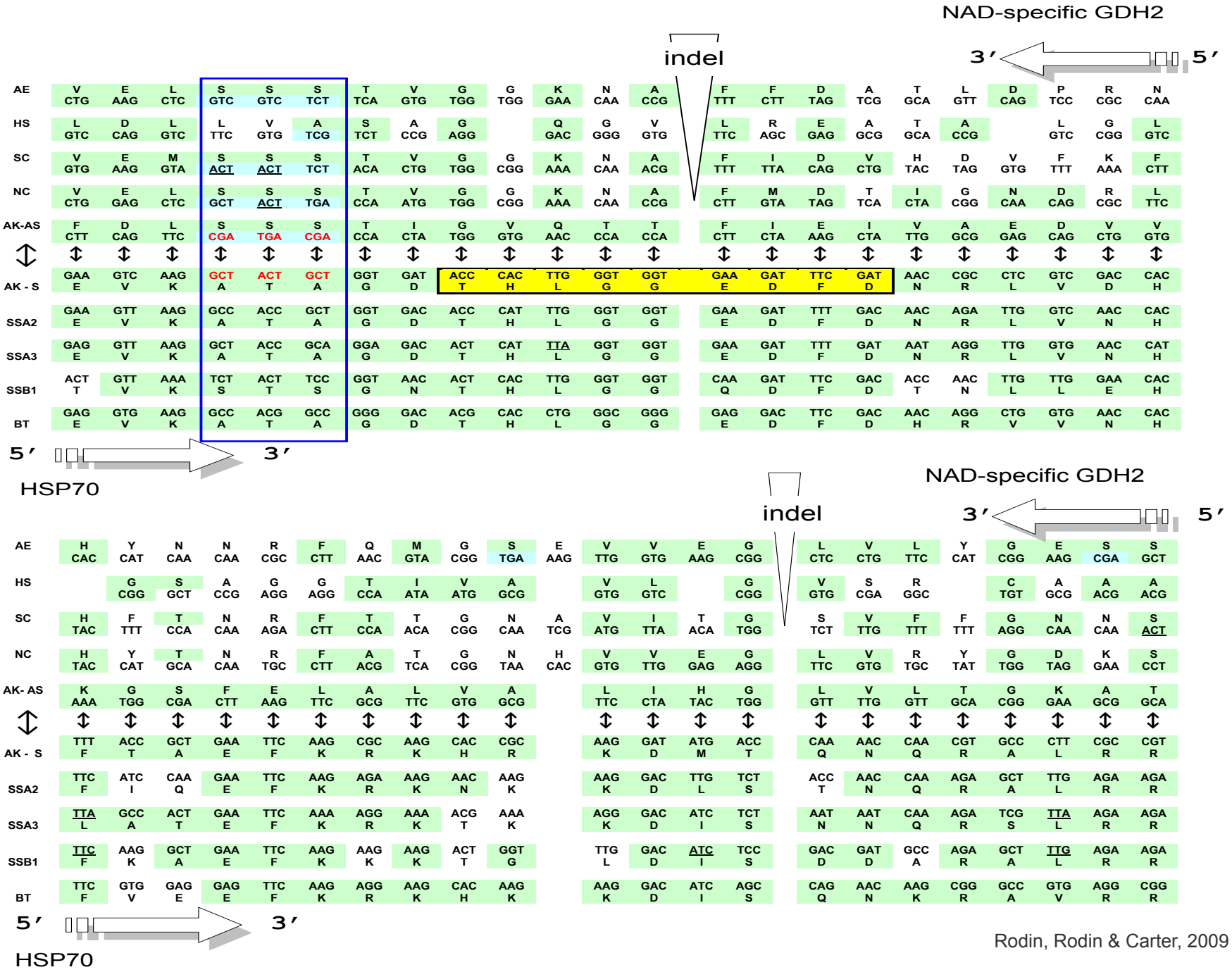
RNA world was more strand-symmetric than the subsequent bilingual world of nucleic acids (RNA, DNA) and proteins

DOUBLE STRAND CODING



Erroneous RNA replication imposed strong limits on the genome size of ribocytes and, therefore, simultaneous recruitment by primitive translation of complementary codons and, accordingly, tRNA pairs with complementary anticodons was an advantage...

The genetic code itself and its two main adaptors, tRNAs and AARSs, might still retain the signatures of this fundamental complementary symmetry (*Eigen & Schuster, 1979; Rodin et al., 1993-2007*).

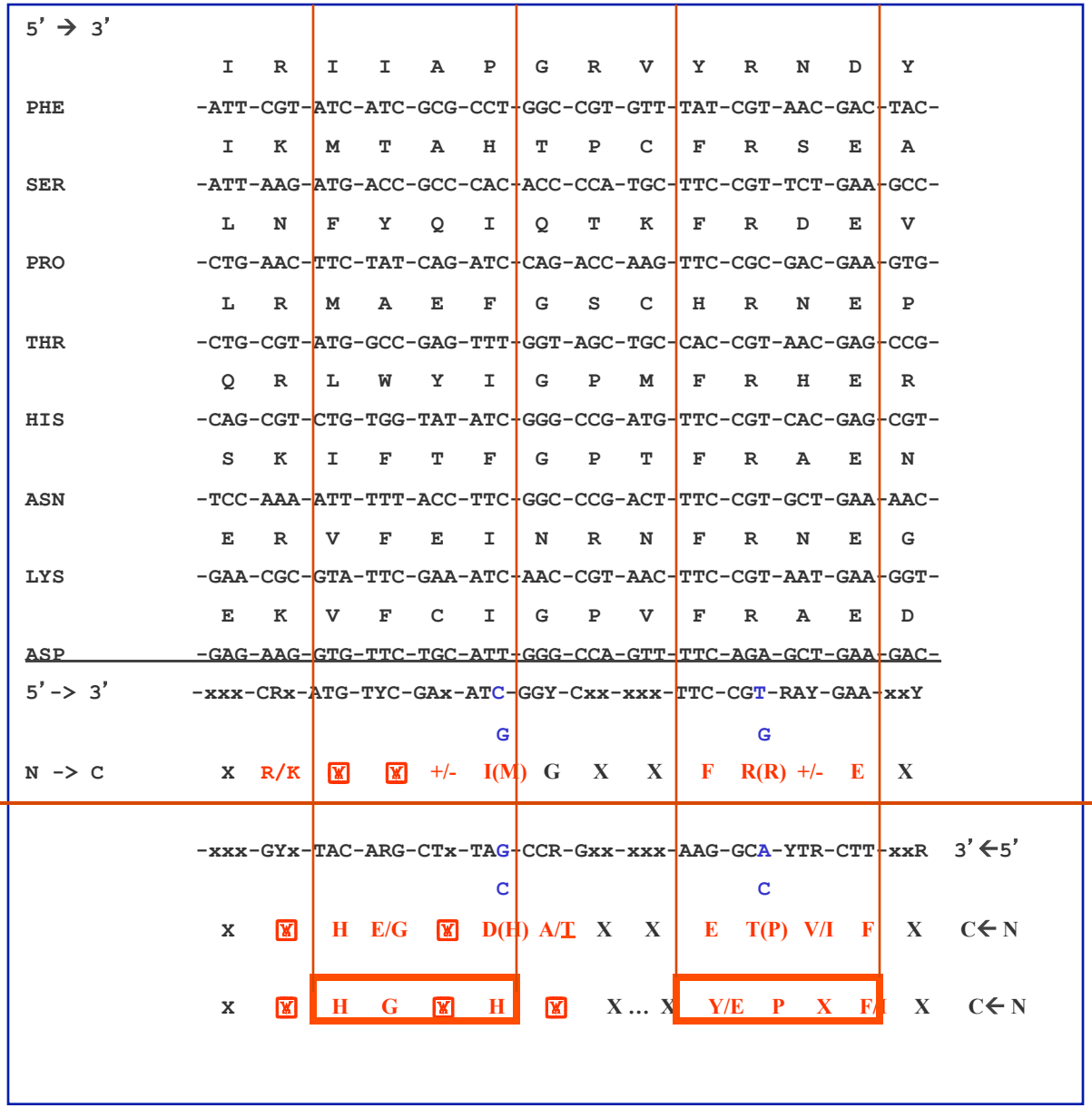


Rodin, Rodin & Carter, 2009

(Rodin & Ohno, 1995, OLEB 25, 565-589)

Class I, PxxxHIGH

Class II, Motif 2

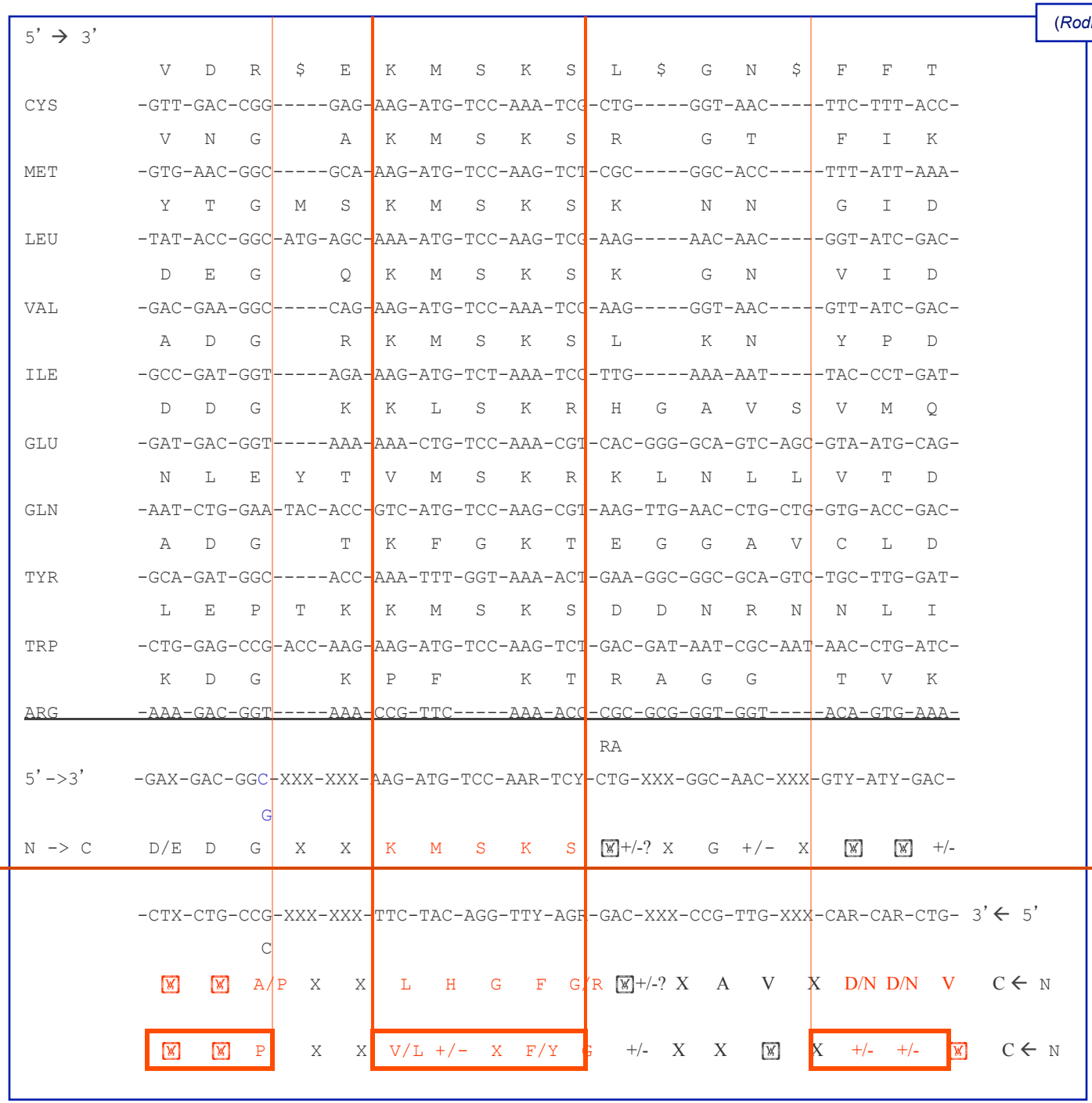


(Rodin & Ohno, 1995, OLEB 25, 565-589)

Class II, Motif 2

Class I, Pxxx HIGH

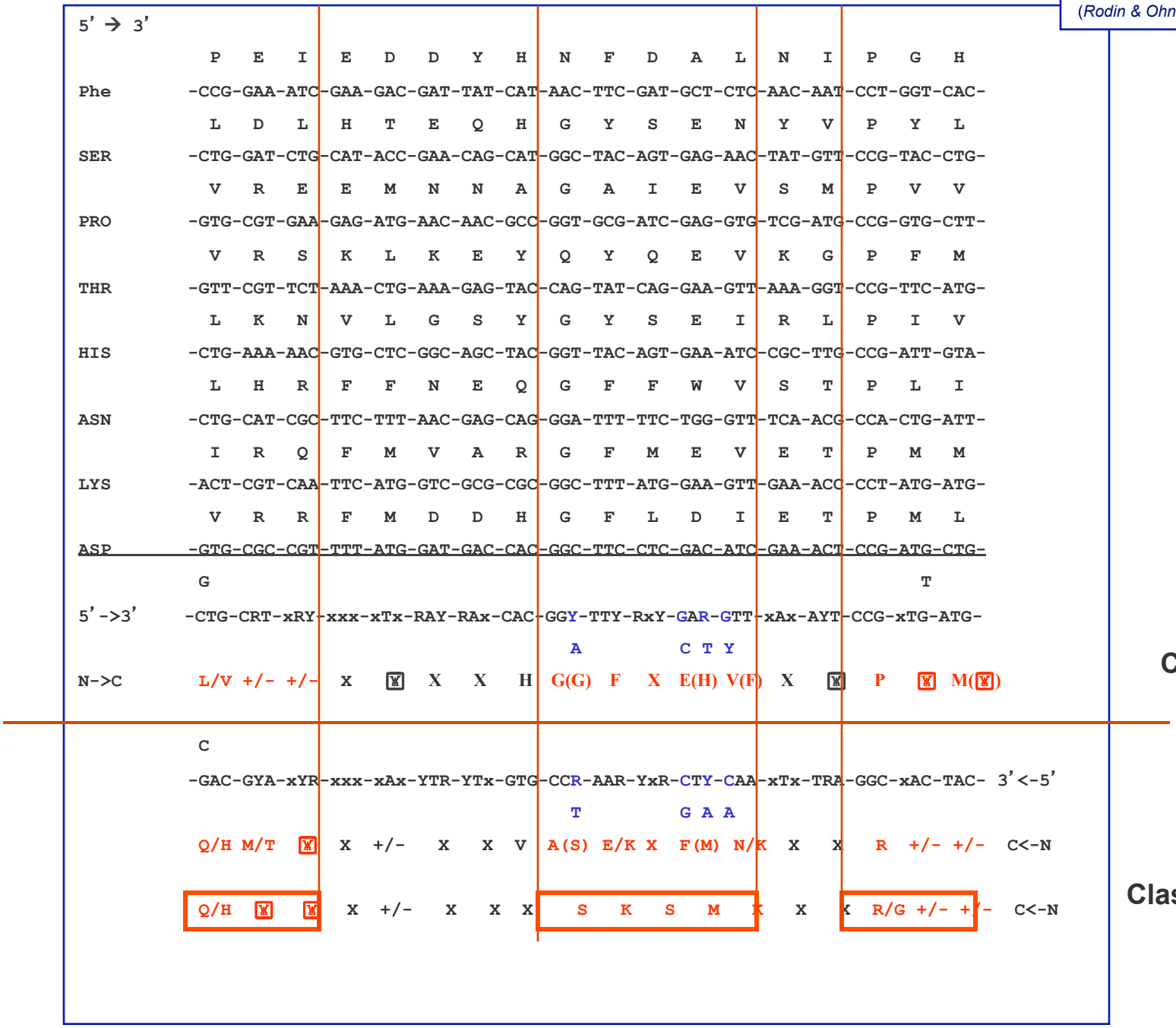
(Rodin & Ohno, 1995, OLEB 25, 565-589)



Class I, …KMSKS…

Class II, Motif 1

(Rodin & Ohno, 1995, OLEB 25, 565-589)

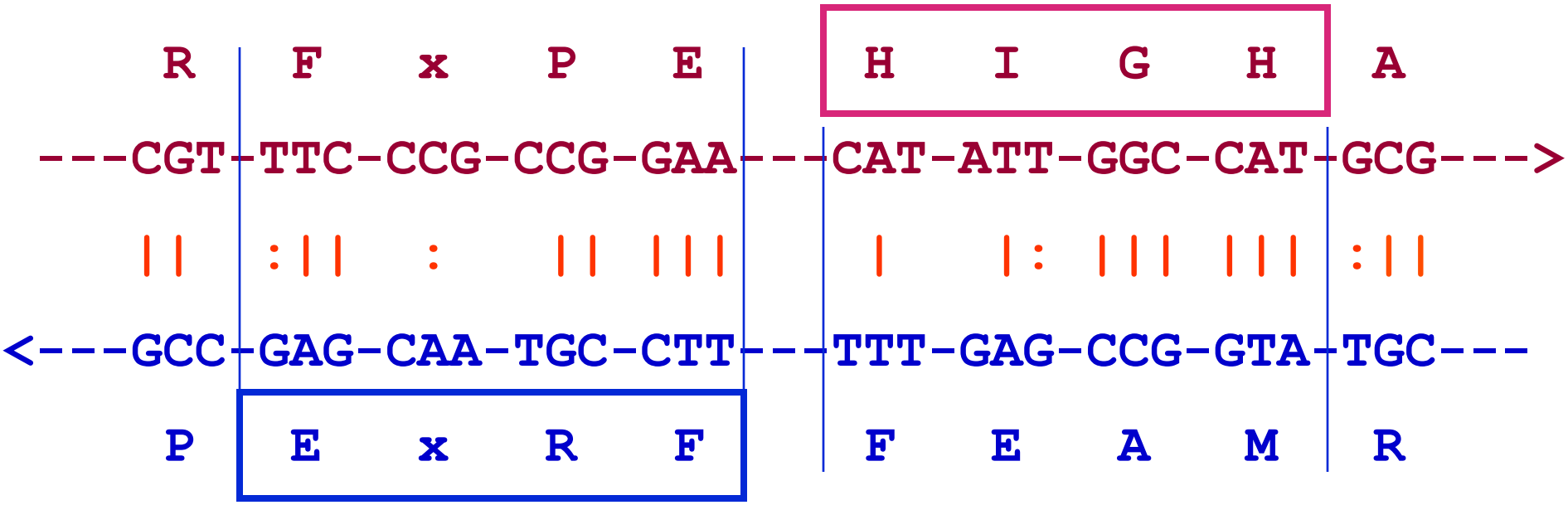


Class II, Motif 1

Class I, … KMSKS …

GlnRS

5’→ 3’



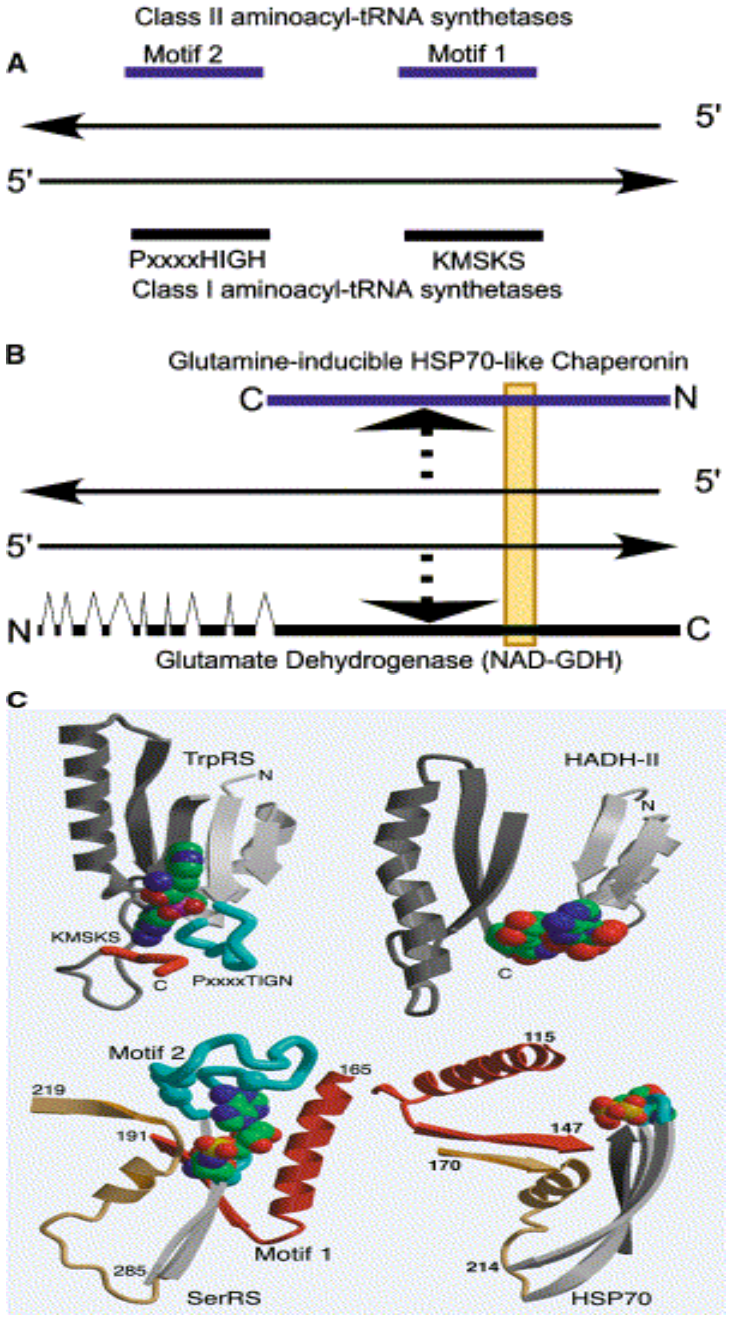
3’← 5’

ThrRS

Carter & Duax, Mol. Cell, 2002

In *Achyla klebsiana*

- **the *GDH* gene encodes a glutamate dehydrogenase on one strand and… (in frame) a heat shock protein HSP70 on the opposite strand**
- **These two proteins are homologous to the Class I and II AARS.**



Carter & Duax, 2002, Mol. Cell, 10, 705-708

Sense-Antisense Relationships and the aaRS Class Distinction (A) Antisense coding of class I (PxxxxHIGH; KMSKS) and class II (motifs 1 and 2) aaRS catalytic motifs (Rodin & Ohno, 1995).

(B) Contemporary proteins coded by in-frame, antisense sequences (LeJohn et al, 1994b). The beige box identifies sequences involved in structural superpositions with class I and class II aaRS.

(C) Nucleotide binding sites in models of the two contemporary sense-antisense proteins (right) and corresponding fragments of classes I and II aaRS (left). Superimposed fragments (CDSFIT [CCP4, 1991]) are light gray; aaRS ATP binding signatures are cyan (motif 2 and TIGN, the TrpRS variant of HIGH) and red (motif 1 and KMSKS). The class IIa TxE signature that orients the α -amino group is in the gold-colored turn connecting the β strand and helix. α

Carter & Duax, 2002, Mol. Cell, 10, 705 - 708

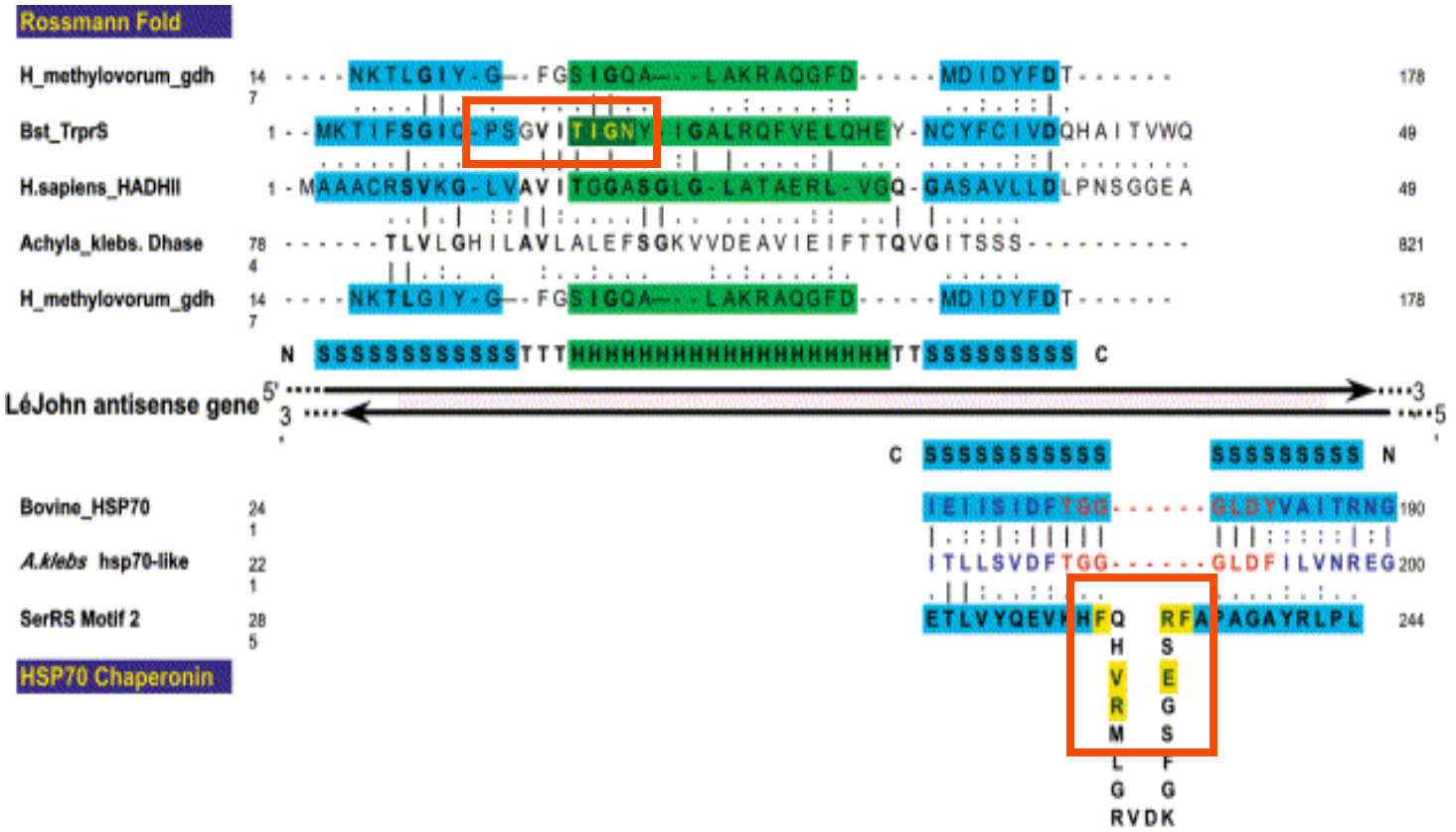


Figure 2. The *A. klebsiana* Sense-Antisense Gene Region Highlighted in Figure 1C Secondary structures are blue for beta strand and green for alpha helix. The DLGGGT HSP70 signature is highlighted by red letters. TIGN and motif 2 sequences are dark green and yellow, respectively. Alignments were performed using EMBOSS (*Rice et al., 2000*) and CLUSTALX (*Thompson et al., 1997*).

The Rodin-Ohno model would require that the two superimposed fragments align opposite one another. The superimposed fragments align closely, within the same 5% of the 650 amino acid sense-antisense coding region (Fig. 2). The 35 amino acid offset of PxxxxHIGH from motif 2 in Fig.2 does not rule out an ancestral sense-antisense relationship between them. Carboxylate clusters apparently migrated comparable distances by mutation and selection as tropomyosin adapted to filamentous actin *(McLachlan & Stewart, 1976*).

"Structures serving a complex function today arose first to serve a simpler one"

(Maynard-Smith & Szathmary "The Major Transitions in Evolution" 1995)

Primary function(s) of p-aaRSs might be much simpler than specific aminoacylation of tRNAs.

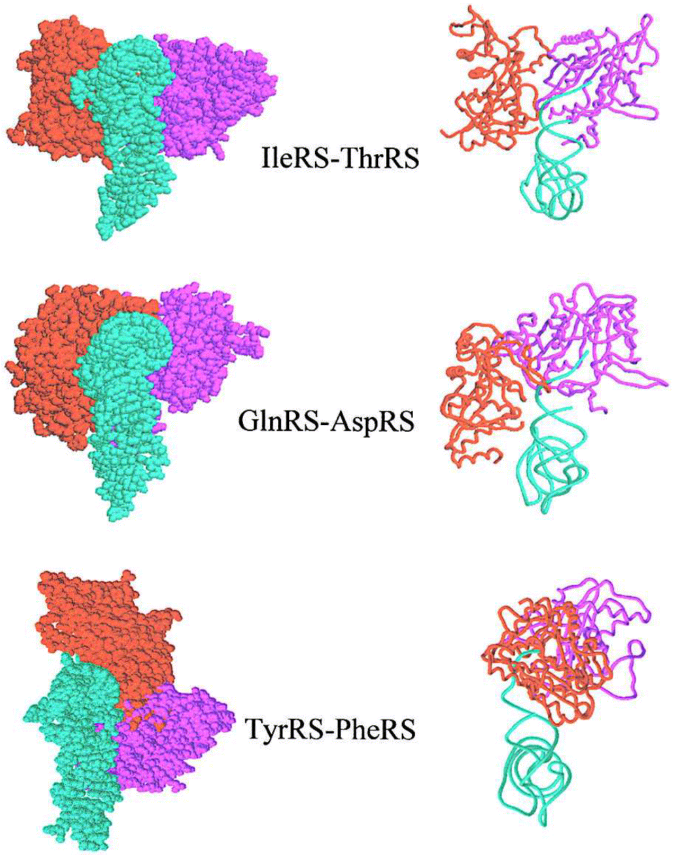
two "complementarily" folded conformations, binding to the opposite sides of tRNA acceptor double helix…

certain other peculiarities revealed by further division of p-aaRSs into subclasses

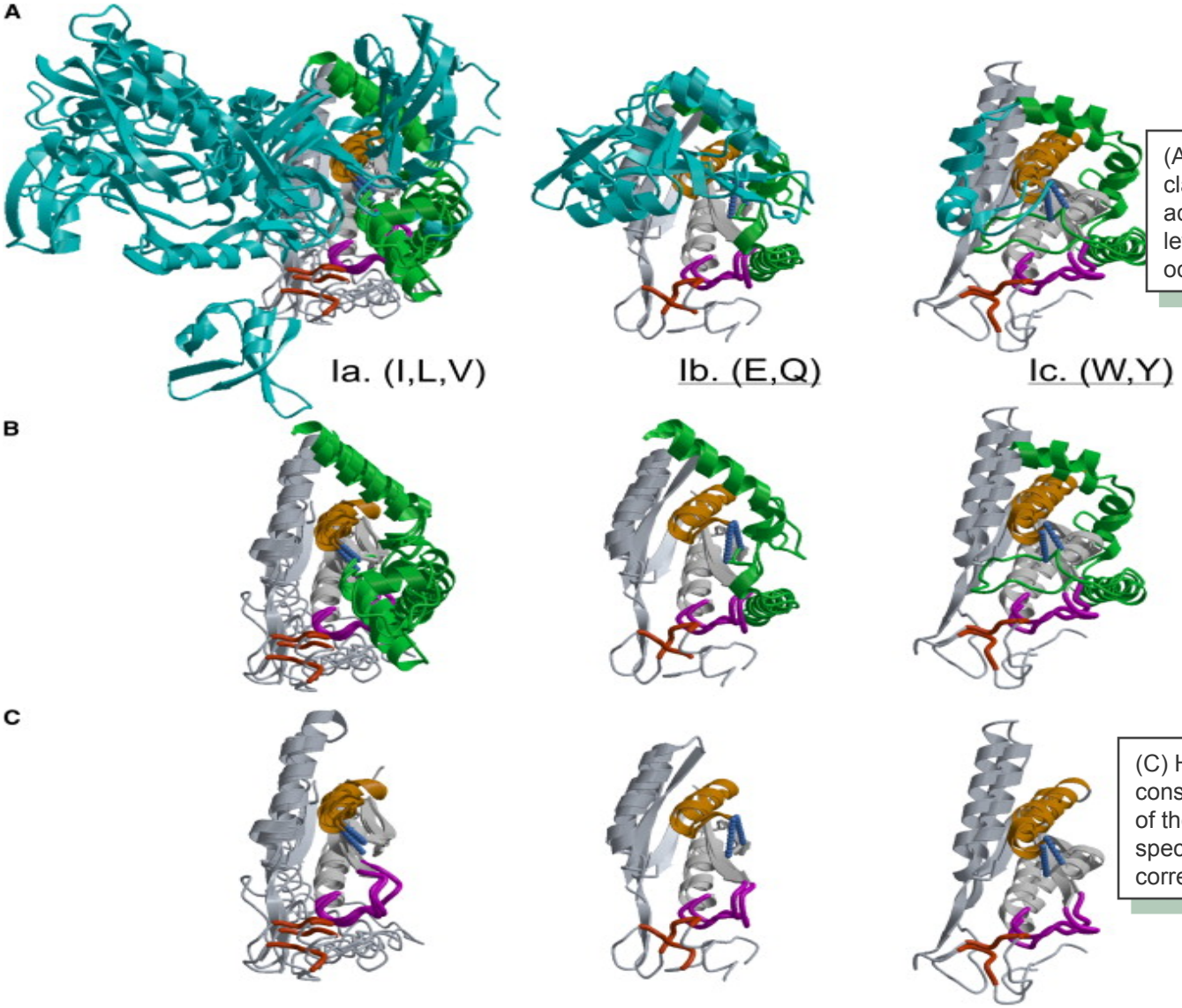
⇩

ancestral pairs of p-aaRSs from complementary classes acted as "chaperones", covering and protecting the acceptor end of tRNAs

(Ribas de Pouplana & Schimmel, 2001)



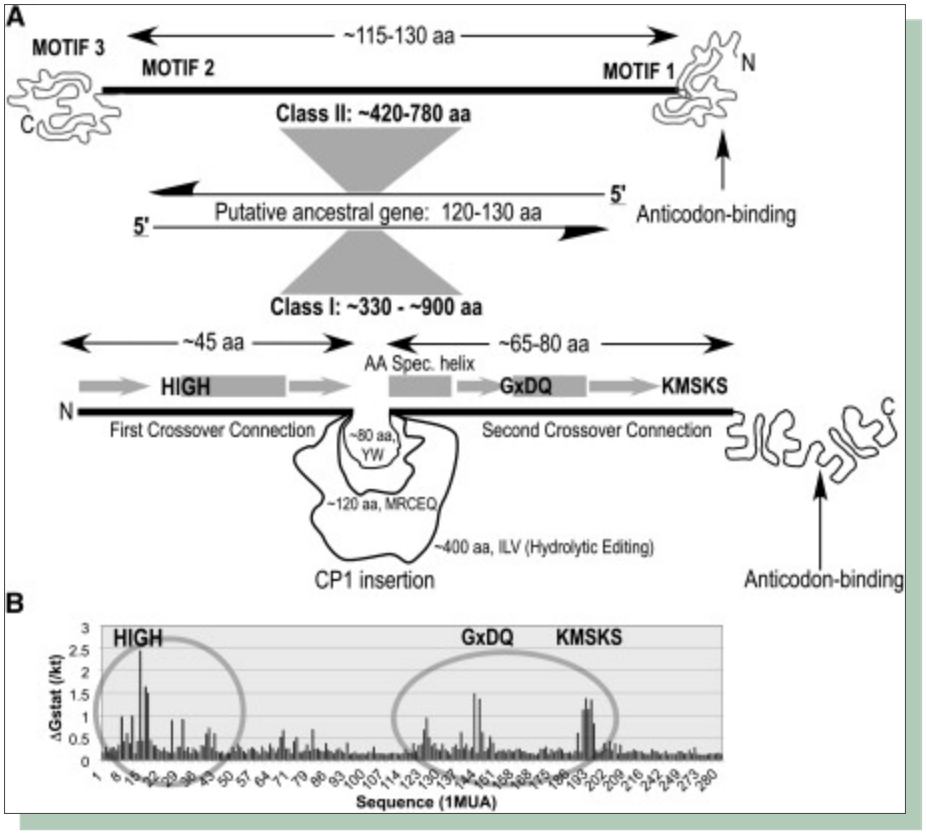
Hierarchical Mosaic Structure in Class I aaRS Catalytic Domains *(Pham et al., Mol. Cell, 2007)*



(A) Catalytic domains of seven of the ten class I aaRS families, superimposed according to subgroups A, B, and C from left to right. Adaptive radiation (cyan) occurs almost exclusively in CP1.

(B) Portions of the CP1 peptide conserved in all class I aaRS (green).

(C) Highly conserved core domain consisting of the two crossover connections of the Rossmann fold (gray) with the specificity-determining α helix (orange), corresponding to the bold lines in Fig. 1A.

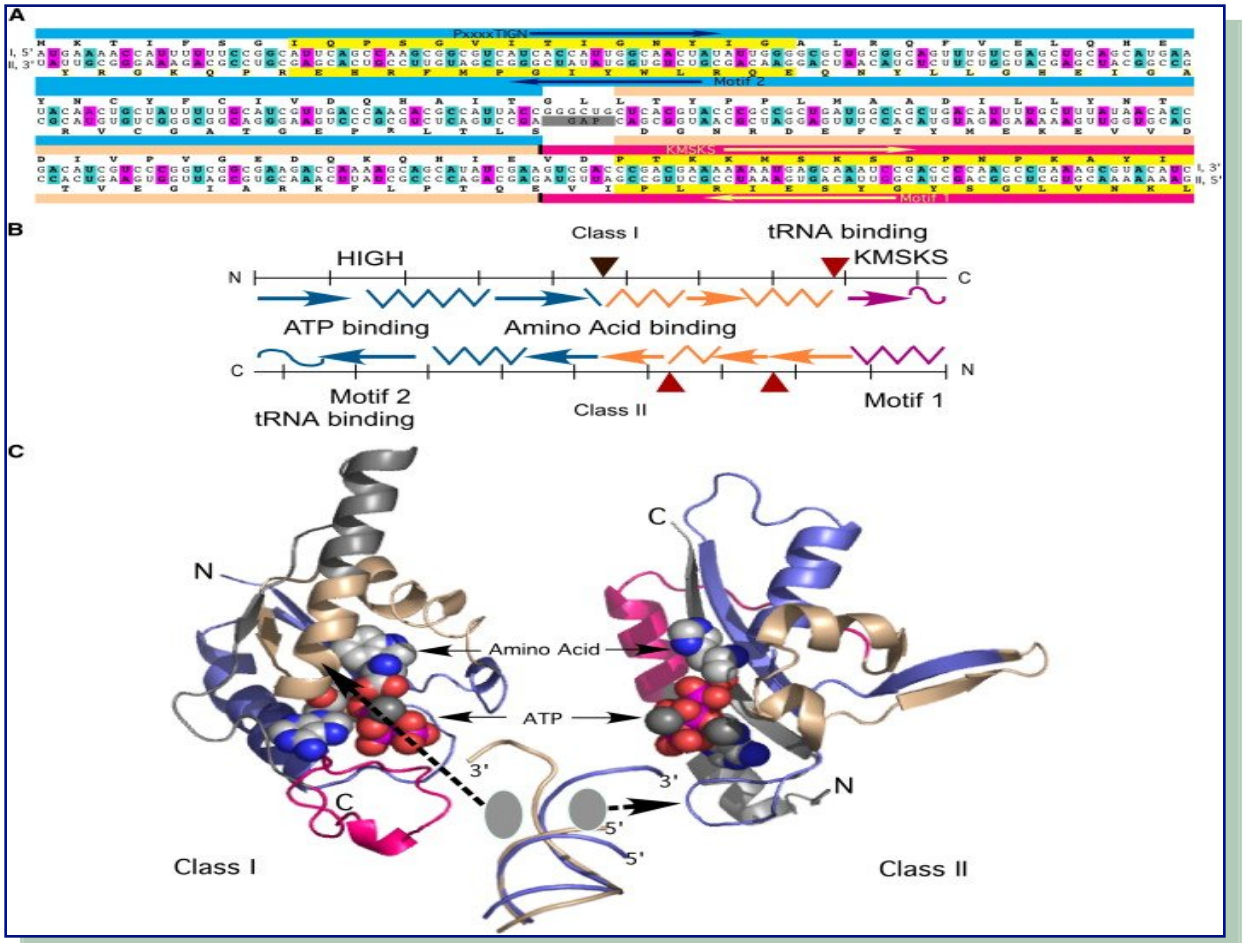


Pham et al., Mol Cell, 2007

aaRS Domain Organization and Sequence Conservation and the Sense/Antisense Coding Hypothesis (Rodin and Ohno, 1995; Rodin and Rodin, 2006b)

(A) Class II core catalytic domains formed between motifs 1 and 2 are generally only 100–130 amino acids long. Conserved class I catalytic sites (bold lines and gray secondary structures) are invariably interrupted by insertions, whose lengths vary from 80 amino acids in TrpRS to 400 amino acids in class Ia aaRSs specific for Ile, Leu, and Val and which are known as CP1 (Burbaum et al., 1990). Idiosyncratic anticodon binding domains occur at the C termini in class I and usually at the N termini in class II.

(B) Sequence conservation across class I aaRS MSAs. The ordinate, ΔG_{stat}, measures the log-likelihood associated with the frequencies of amino acids in each position of the MSA (Suel et al., 2003). High values indicate sequence conservation. The two gray ovals correspond to the two halves of the Rossmann fold (bold lines in [A]). Characters in boldface are active-site catalytic residues.



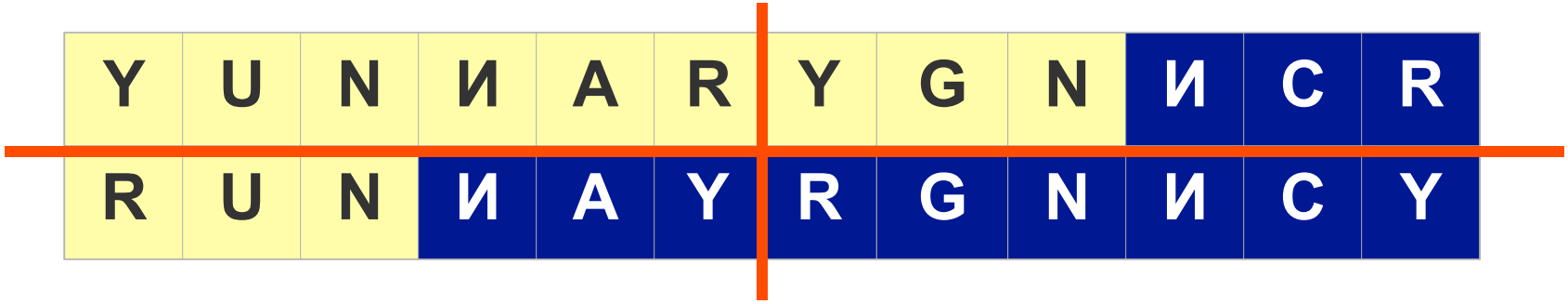
Pham et al., Mol Cell, 2007

Ancestral Sense/Antisense "Gene" for Minimal Catalytic Domains Coding that for Class I TrpRS Opposite that for Class II HisRS

(A) Nucleotide sequences for the catalytic domains drawn from the genes for *B. stearothermophilus* TrpRS (top strand) and *E. coli* HisRS (bottom strand) were aligned using GAP (Accelrys, 2006). The alignment matched both pairs of signature sequences to each other (yellow color and arrows), as expected from the Rodin-Ohno hypothesis.

(B) (B) Secondary structures encoded by the 96 amino acids in (A), including functionally important residues for ATP, amino acid, and tRNA binding.

(C) (C) Structures of the TrpRS and HisRS MCDs are based on the respective crystal structures and superimposed using the β strand loop regions culminating in the TIGN and motif 2 signatures, resulting in antiparallel orientations of the two active sites. A superimposed pair of class I and II tRNA acceptor stems is indicated, with dashed arrows to primary binding surfaces in the two classes. As noted by Ribas de Poipuplana and Schimmel (2001b), the tRNA binding sites approach from opposite directions (upper left in TrpRS, lower right in HisRS).



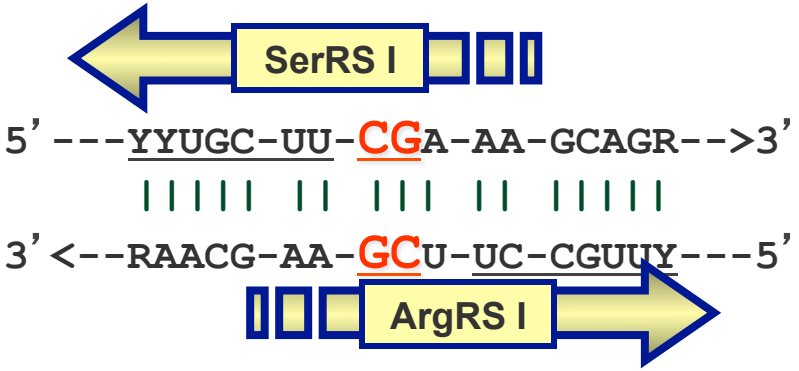
The 10:10 ratio

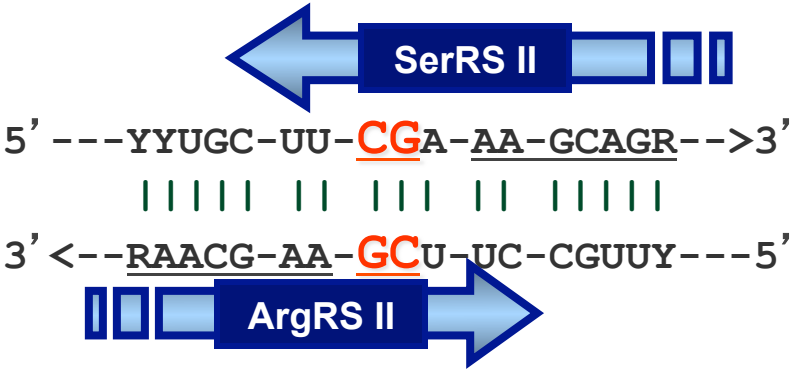
This “yin-yang” split divides all codons in two equal (32 + 32) groups, corresponding to the I and II classes of aaRSs. What provides for this equality is the double-strand coding. This does not necessarily imply that the amino acids must be equally represented by the two aaRS classes --- only two roughly commensurable groups are expected. Indeed, the actual ratio is 9.5 : 10.5 (if one takes into account the “Janusian” lysine).

The binary sub-code of aminoacylation makes the 10:10 ratio looking much less miraculous.

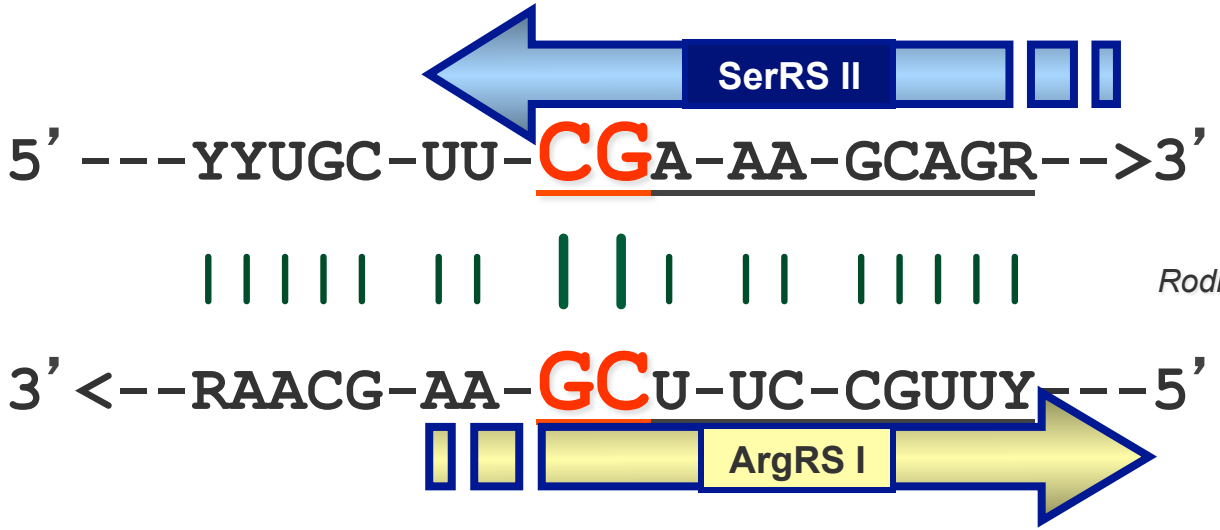
Is this “yin-yang” pattern of tRNA recognition perfectly symmetric with regard to the two modes – from major and minor groove sides?

HIGH RISK of erroneous aminoacylation

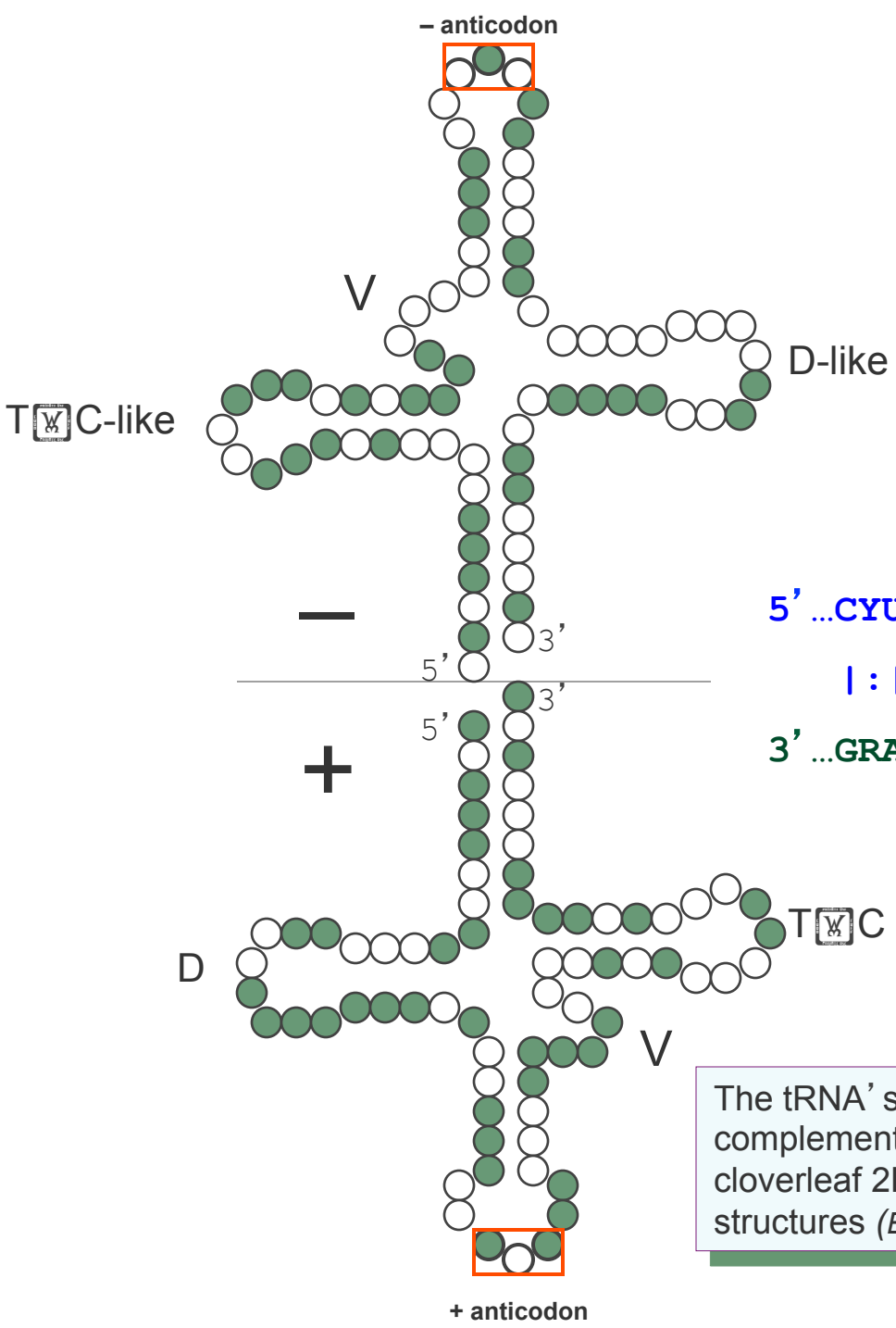




LOW RISK of erroneous aminoacylation



Rodin, Ohno, Rodin, 1993



ALA

→

5’…CYUGC-UU-CGC-AA-GCAGG…3’

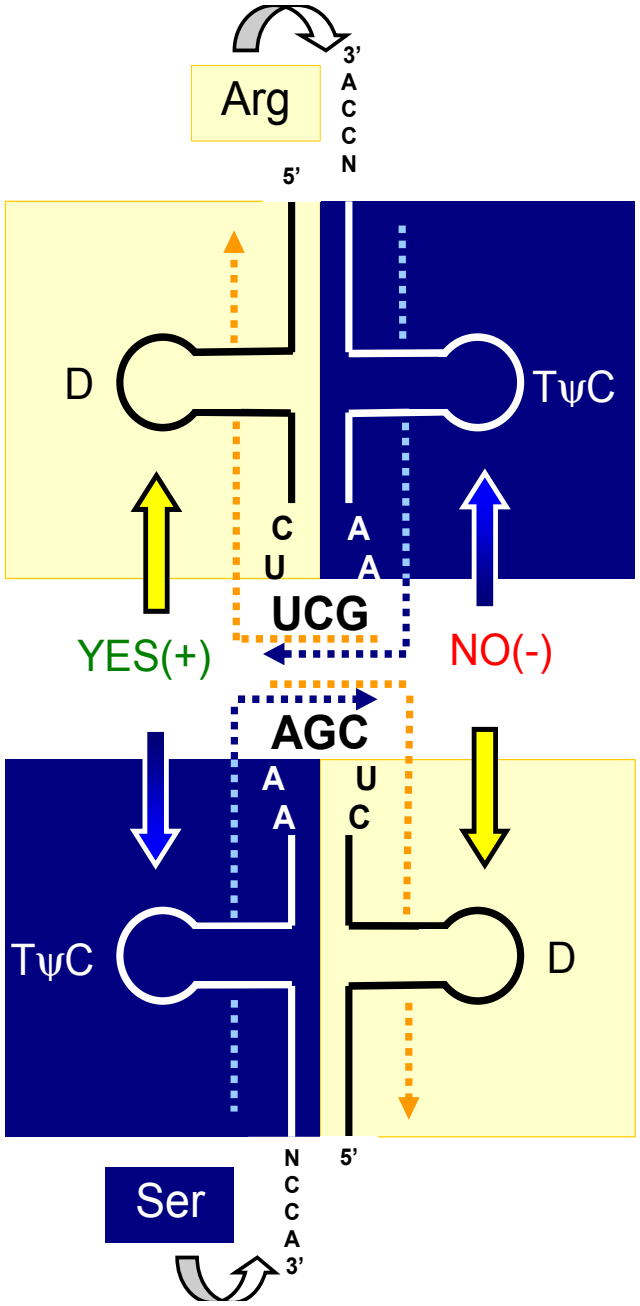
I:III II III I* III::

3’…GRACG-AG-GCG-UC-CGUUU…5’

←

ARG

The tRNA’s mirror replica with the complementary anticodon shows the same cloverleaf 2D and the similar L-shaped 3D structures *(Eigen & Schuster, 1979).*



Arg (UCG) vs. (CGA) Ser

5' x 5' 3' 5' 5'.....CU **UCG** AA.....3' x 3' 5' 5'.....CU **CGA** AA.....3' −

5' x 3' 3' 5' 5'.....CU **UCG** AA.....3' x 3' 5' 5'.....CU **CGA** AA.....3' +

3' x 3' 3' 5' 5'.....CU **UCG** AA.....3' x 3' 5' 5'.....CU **CGA** AA.....3' −

3' x 5' 3' 5' 5'.....CU **UCG** AA.....3' x 3' 5' 5'.....CU **CGA** AA.....3' −

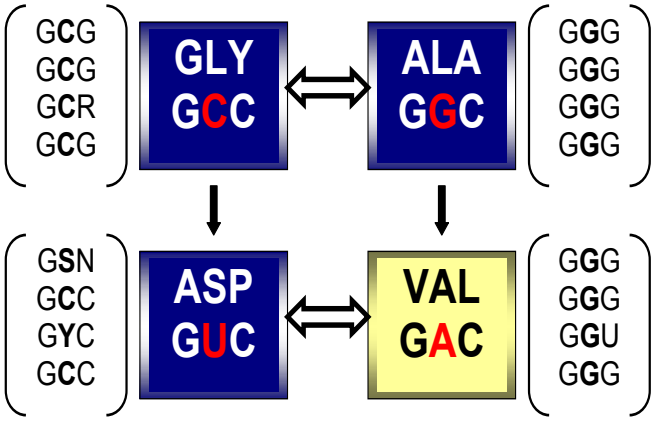
⇒ r-aaRS I
⇒ r-aaRS II

5' - cU XYZ AN - 3'

The yin/yang-like internal sub code for two modes of tRNA aminoacylation revealed by the complementary re-arrangement of the code table minimizes the risk of confusion of tRNAs with complementary anticodons.

Rodin & Rodin, 2006, 2008

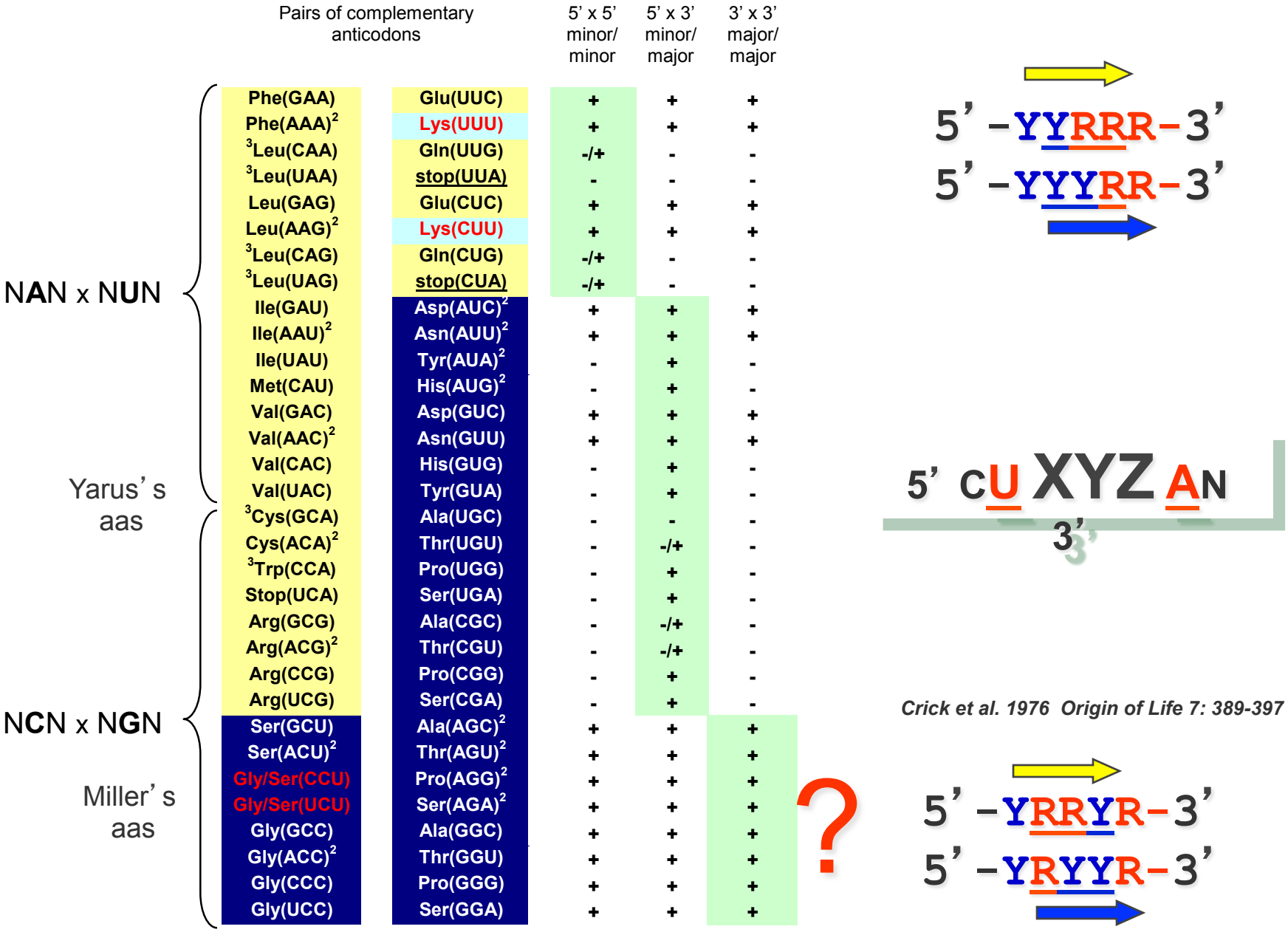
Stanley Miller's aas



Rodin & Rodin, 2008 Heredity, in press

	"SENSE" old	"SENSE" new		"ANTISENSE" old	"ANTISENSE" new
aa	ALA	VAL		GLY	ASP
mRNA	gCc -> gUc		vs.	gGc -> gAc	
tRNA	gGc	gAc		gCc	gUc
r-aaRS	gCc	gUc		gGc	gAc

Pleiotropic negative effect


Pairs of complementary anticodons
5' x 5' minor/ minor
5' x 3' minor/ major
3' x 3' major/ major
NAN x NUN
Yarus' s aas
NCN x NGN
Miller' s aas
Phe(GAA) Glu(UUC) + + +
Phe(AAA)2 Lys(UUU) + + +
3Leu(CAA) Gln(UUG) -/+ - -
3Leu(UAA) stop(UUA) - - -
Leu(GAG) Glu(CUC) + + +
Leu(AAG)2 Lys(CUU) + + +
3Leu(CAG) Gln(CUG) -/+ - -
3Leu(UAG) stop(CUA) -/+ - -
Ile(GAU) Asp(AUC)2 + + +
Ile(AAU)2 Asn(AUU)2 + + +
Ile(UAU) Tyr(AUA)2 - + -
Met(CAU) His(AUG)2 - + -
Val(GAC) Asp(GUC) + + +
Val(AAC)2 Asn(GUU) + + +
Val(CAC) His(GUG) - + -
Val(UAC) Tyr(GUA) - + -
3Cys(GCA) Ala(UGC) - - -
Cys(ACA)2 Thr(UGU) - -/+ -
3Trp(CCA) Pro(UGG) - + -
Stop(UCA) Ser(UGA) - + -
Arg(GCG) Ala(CGC) - -/+ -
Arg(ACG)2 Thr(CGU) - -/+ -
Arg(CCG) Pro(CGG) - + -
Arg(UCG) Ser(CGA) - + -
Ser(GCU) Ala(AGC)2 + + +
Ser(ACU)2 Thr(AGU)2 + + +
Gly/Ser(CCU) Pro(AGG)2 + + +
Gly/Ser(UCU) Ser(AGA)2 + + +
Gly(GCC) Ala(GGC) + + +
Gly(ACC)2 Thr(GGU) + + +
Gly(CCC) Pro(GGG) + + +
Gly(UCC) Ser(GGA) + + +
?
5' -YYRRR-3'
5' -YYYRR-3'
5' cU XYZ AN 3'
Crick et al. 1976 Origin of Life 7: 389-397
5' -YRRYR-3'
5' -YRYYR-3'
Rodin & Rodin, 2008

(Rodin, Rodin & Ohno, 1996)

Table 1. Correlated complementarity of the second bases in anticodon and acceptor

Class I × Class I pairs	Class I × Class II pairs	Class II × Class II pairs
Leu (**C**, TAA) ⇔ Gln (**G**, TTA*)	Phe (**C**, GAA) ⇔ Glu (**C**, TTC)	Phe (**C**, AAA) ⇔ Lys (**R**, TTT)
Leu (**C**, CAA) ⇔ Gln (**G**, TTG)	Leu (**G**, AAG) ⇔ Lys (**C**, CTT)	Ser1 (**G**, GGA) ⇔ Gly (**C**, TCC)
Leu (**G**, GAG) ⇔ Glu (**C**, CTC)	Ile (**G**, AAT) ⇔ Asn (**C**, ATT)	Pro (**G**, GGG) ⇔ Gly (**C**, CCC)
Leu (**C**, TAG) ⇔ Gln (**G**, CTA*)	Ile (**G**, GAT) ⇔ Asp (**C**, ATC)	Thr (**C**, GGT) ⇔ Ser2 (**G**, GCT)
Leu (**Y**, CAG) ⇔ Gln (**G**, CTG)	Met (**G**, CAT) ⇔ His (**Y**, GTG)	Thr (**C**, GGT) ⇔ Gly (**C**, GCC)
Ile (**G**, TAT) ⇔ Tyr (**S**, GTA)	Val (**G**, AAC) ⇔ Asn (**C**, GTT)	Ala (**G**, AGC) ⇔ Ser2 (**G**, GCT)
Val (**T**, TAC) ⇔ Tyr (**S**, GTA)	Val (**G**, GAC) ⇔ Asp (**C**, GTC)	Ala (**G**, GGC) ⇔ Gly (**C**, GCC)
	Val (**G**, CAC) ⇔ His (**Y**, GTG)	
	Ser1 (**G**, AGA) ⇔ Arg (**C**, TCT)	
	Ser1 (**G**, TGA) ⇔ Trp (**S**, TCA*)	
	Ser1 (**G**, CGA) ⇔ Arg (**C**, TCG)	
	Pro (**G**, RGG) ⇔ Arg (**C**, CCT)	
	Pro (**G**, TGG) ⇔ Trp (**S**, CCA)	
	Pro (**G**, CGG) ⇔ Arg (**C**, CCG)	
	Thr (**C**, TGT) ⇔ Cys (**S**, GCA)	
	Thr (**C**, CGT) ⇔ Arg (**G**, ACG)	
	Ala (**G**, TGC) ⇔ Cys (**S**, GCA)	
	Ala (**G**, CGC) ⇔ Arg (**T**, GCG)	

Shown are pairs of complementary anticodons and the respective 2nd bases in the 5′ acceptor strand of the corresponding consensus tRNA genes. To use all 32 pairs, the tRNA genes, not tRNA themselves, are presented in the table, and three "nonsense" anticodons (marked by asterisks) are assigned to Gln (TTA, CTA) and Trp (TCA) according to ref. 25. There are two unlinked groups of such pairs for serine designated Ser1 and Ser2, respectively. The 2nd bases in the acceptor and anticodon are in boldface type. Underlined are the three exceptions when the complementarity of the anticodons is not accompanied by the complementarity of the 2nd bases in the 5′ strand of the corresponding acceptor helices.

? N (I x I, II x II)  N (I x II)

Dual complementarity of second bases in separate organisms and consensus tRNAs representing main kingdoms *(Rodin & Ohno, 1997)*

Organism or group	No. of pairs of tRNAs with complementary anticodons	No. of pairs with complementary second bases in the acceptors	No. of pairs with noncomplementary second bases in the acceptor	
E. coli	32	24	8	P ≈ 0.008
H. volcanii	29	24	5	
S. cerevisiae	24	20	4	P ≈ 0.04
Chloroplast	26	19	7	
Cytoplasm of plants	20	16	4	p = 0.008
Cytoplasm of animals	27	18	9	P ≈ 0.04
Mitochondria of fungi	18	12	6	P ≈ 016
Mitochondria of plants	17	12	5	
Mitochondria of animals	17	9	8	
Pooled data	210	154	56	p = 0.008
Common consensus	32	29	3	p < 0.00001

MAIN RESULT: Dual complementarity is shown by ancestral/ consensus tRNA pairs with completely complementary anticodons and is not shown by tRNA pairs in which only the 2nd bases of anticodons are complementary

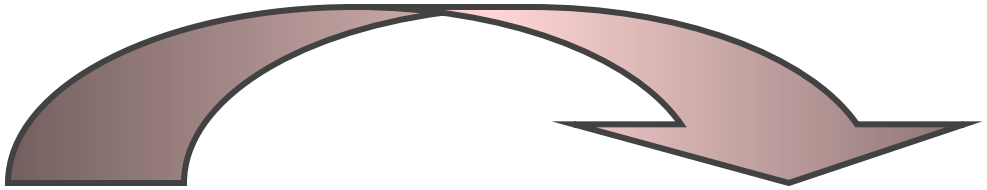
The following corollaries immediately follow:

- **By the time ancestral tRNAs gained the dual complementarity, the 3-letter translation frame has already been in use.**

- **The very phenomenon of dual complementarity is possible largely because the new tRNAs entered primitive translation in pairs with complementary anticodons.**

The updated genomic tRNA compilation of 8,246 tRNA gene sequences (instead of 1,268 ones available 10 years ago)

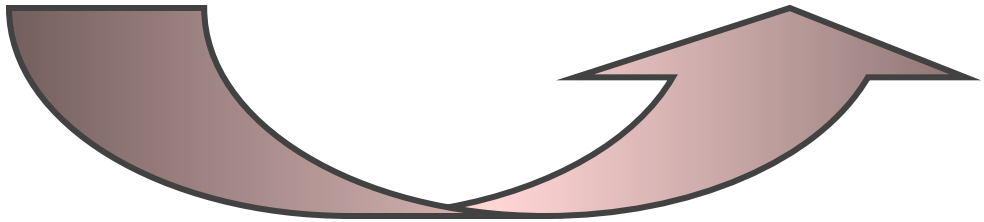
A closer (in *re* dual complementarity) comparison of pairs of ancestral tRNAs with completely complementary anticodons with those having only central complementary bases revealed the following:



5’	U		C		A		G		3’
	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
U	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	*Stop*	UGA	*Stop*	A
	UUG	Leu	UCG	Ser	UAG	*Stop*	UGG	Trp	G
	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
C	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
A	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
G	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

I class AARS

II class AARS



	PRO GGG	vs.	CYS GCA	TRP CCA	ARG ACG	ARG GCG	ARG TCG	ARG CCG	SER GCT	ARG TCT	ARG CCT	GLY GCC	GLY TCC	GLY CCC
Eubacteria	G		G	G/C	C	T	C	C	G	C/T	C/T	C	C	C
Archaebacteria	G		C	G	-	C	G	G	C	G	G	C	C	C
Thermobacteria*	G	vs.	G	G/C	G/C	T	-	G	G	C	C	C	C	C
Eukaryotes (lower)	G		G	A/G	G/T	-	G/A	G/C	A/T	C/T	C/T	C	C	C
Eukaryotes (higher)	G		G	G	G	C	G/A	G/A	A/T	C	C	C	C	C
Common ancestor	G	vs.	G	G	G/C	T/C	G	G	G/A	C	C	C	C	C
Complementarity					?	+				+	+	+	+	+

Dual Complementarity:

Pairs of GGG vs. anticodons complementary at all three positions DC = 4/4 = 1

Pairs of GGG vs. anticodons complementary at the 2nd position DC = 2/7 = 0.29

ANTICODON PAIRS

2nd base pair: G vs.C or A vs.U

W-C canonical pairs

5’ ---G-G(A)-U---3’

3’ ---C-C(U)-A---5’

COMPLEMENTARITY at the 2nd position of the acceptor stem

	Min	Max
$DC_{(A-U)}$	0.87	0.93
$DC_{(G-C)}$	0.69	0.875

Flanking wobbling pairs

5’ ---G-G(A)-C*---3’

3’ ---U-C(U)-A---5’

$DC_{(A-U)}$	0.86	0.89
$DC_{(G-C)}$	0.68	0.91

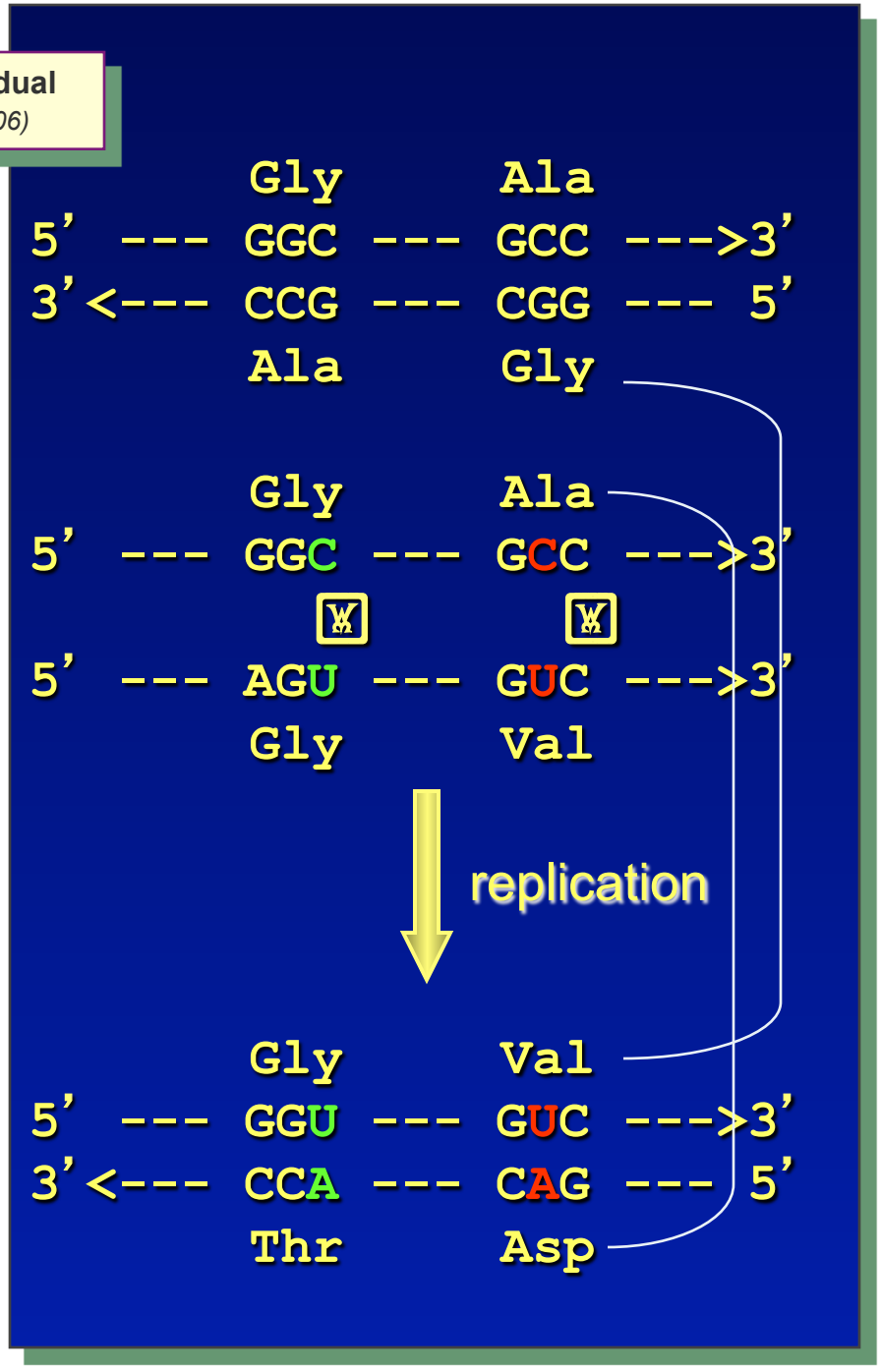
Flanking mispairings

5’ ---R-G(A)-Y---3’

3’ ---R-C(U)-Y---5’

$DC_{(A-U)}$	0.38	0.44
$DC_{(G-C)}$	0.39	0.51

Double strand coding produces dual complementarity *(Rodin & Rodin, 2006)*



Gly → Gly:

D_m= 0

Ala → Thr:

D_m= 0.9

Ala → Val:

D_m= 1.85

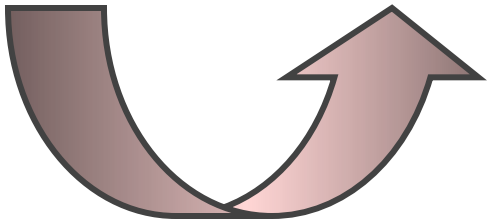
Gly → Asp:

D_m= 2.37

5’	U		C		A		G		3’
	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
U	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	*Stop*	UGA	*Stop*	A
	UUG	Leu	UCG	Ser	UAG	*Stop*	UGG	Trp	G
	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
C	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
A	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
G	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

I class AARS

II class AARS



ANTICODON PAIRS

2nd base pair: G vs. U or A vs.C

```
5’ ---G-G-U---3’
      I I I
3’ ---C-U-A---5’
```

```
5’ ---G-A-U---3’
      I * I
3’ ---C-C-A---5’
```

```
5’ ---R-G(A)-Y---3’
      I   *
3’ ---R-U(C)-Y---5’
```

COMPLEMENTARITY at the 2nd position of the acceptor stem

$DC = 0.44 – 0.56$

$DC_{(wobbling)} = 0.45 – 0.52$

$DC = 0.39 – 0.58$

$DC_{(wobbling)} = 0.41 – 0.59$

$DC_{(G\text{-}U)} = 0.43 – 0.48$

$DC_{(A*C)} = 0.42 – 0.51$

Dual complementarity in ancestral tRNA pairs

ANTICODON PAIRS	COMPLEMENTARITY (at the 2nd position of the acceptor stem)
A. 2nd base pair of anticodons is G – C or A – U	
Flanking normal (Watson-Crick) pairings: 5′ ---> 3′ G **G(A)** U | **|** | C **C(U)** A 3′ <--- 5′	min max DC_{A-U} = 0.87 – 0.93 DC_{G-C} = 0.69 – 0.88
Flanking wobbling (G–U or A*C) pairings 5′ ---> 3′ G **G(A)** **C** **:** **|** * U **C(U)** A 3′ <--- 5′	min max DC_{A-U} = 0.86 – 0.89 DC_{G-C} = 0.68 – 0.91
Flanking R * R or Y * Y mispairings: 5′ ---> 3′ R **G(A)** Y * | * R **C(U)** Y 3′ <--- 5′	min max DC_{A-U} = 0.38 – 0.44 DC_{G-C} = 0.39 – 0.51
B. 2nd base pair of anticodons: G -- U or A * C	
Flanking normal or wobbling pairings: 5′--->3′ R **G** Y | **:** | Y **U** R 3′<---5′	min max DC_n = 0.44 – 0.56 DC_w = 0.45 – 0.52
Flanking normal or wobbling pairings: 5′--->3′ R **A** Y | * | Y **C** R 3′<---5′	min max DC_n = 0.39 – 0.58 DC_w = 0.41 – 0.59
Flanking R * R or Y * Y mispairings: 5′ ---> 3′ R **G(A)** Y * **:**(*) * R **U(C)** Y 3′ <--- 5′	min max DC_{G-U} = 0.43 – 0.48 DC_{A*C} = 0.42 – 0.51

DC = 0.86

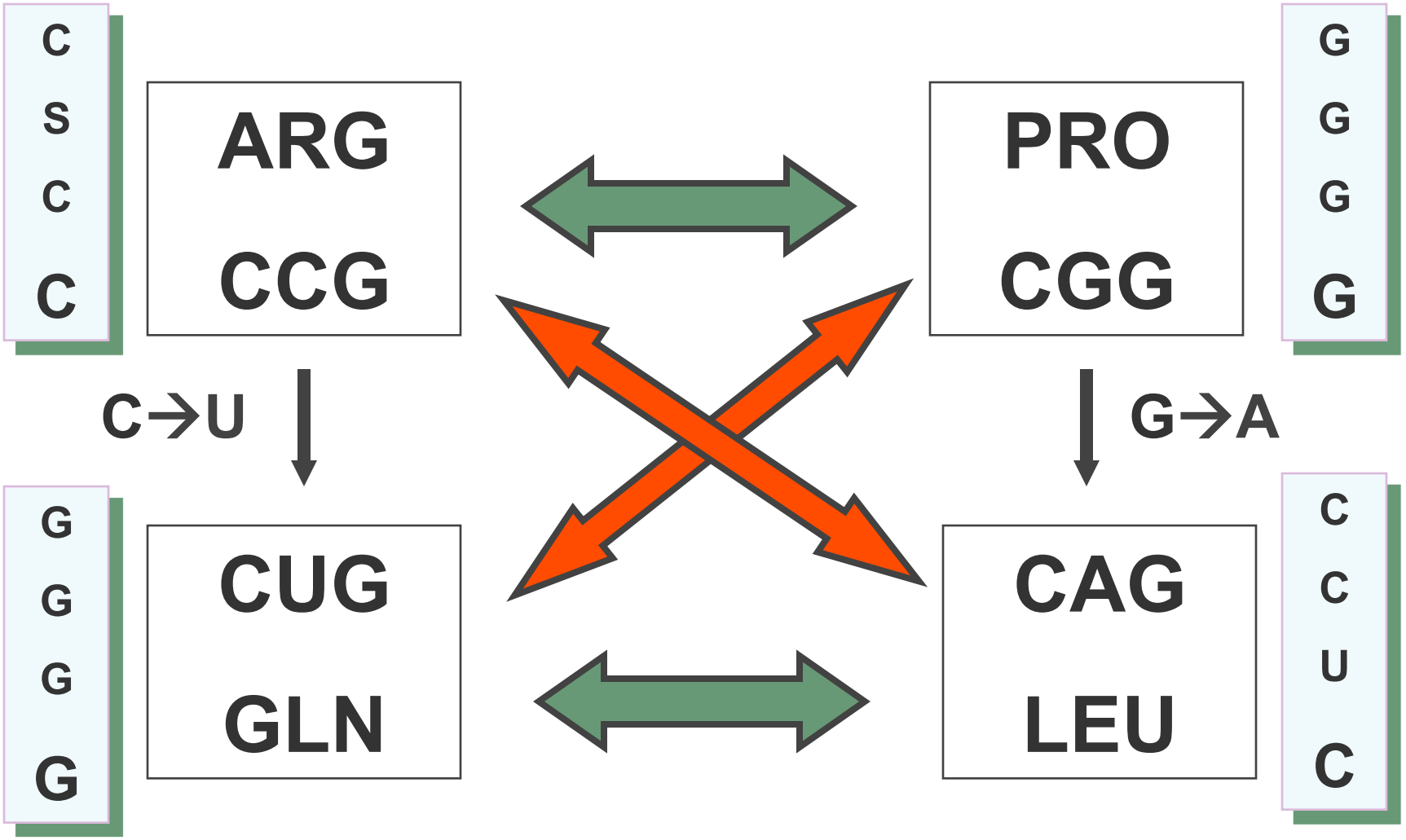
vs.

DC = 0.43

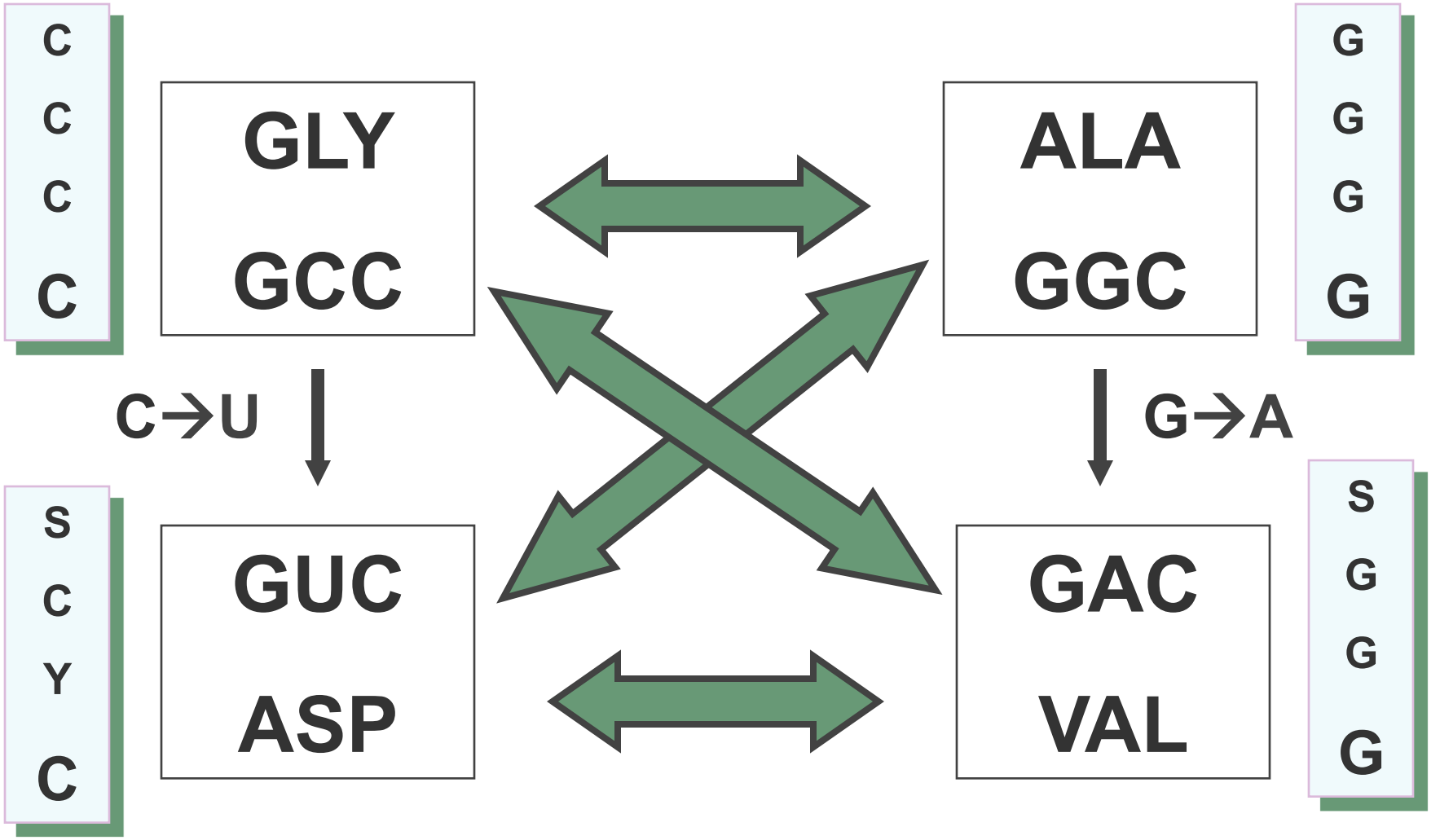
DC = 0.49

vs.

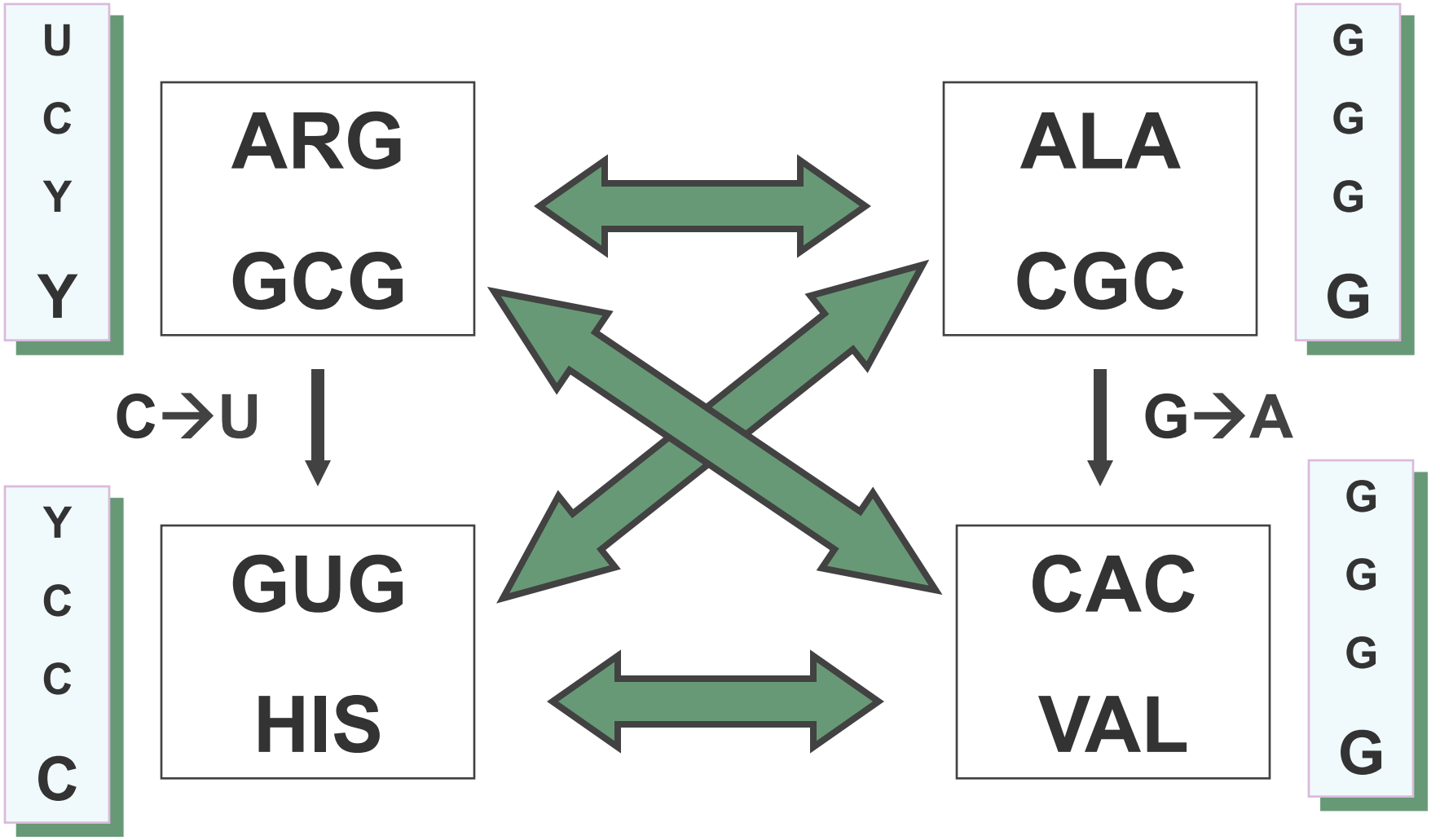
DC = 0.46



A typical example: tRNA pairs with legitimate W-C base pairings at the anticodon 2nd position are also complementary at the 2nd position of their acceptors (green arrows). It it is not true for illegitimate G:U and A*C cases (red arrows). 14 of 16 tetrads are of this type

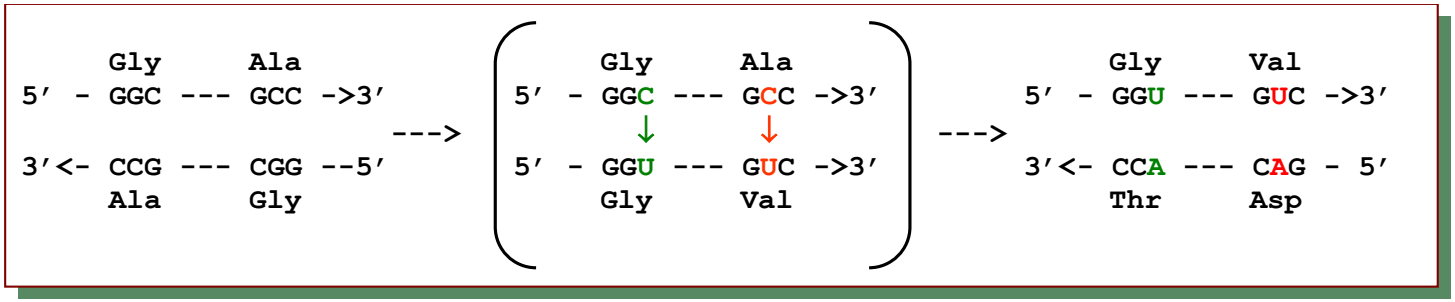


These four amino acids are complementary at the 2nd position of the acceptor stem in all combinations (legitimate W-C and "wobbling" pairs).

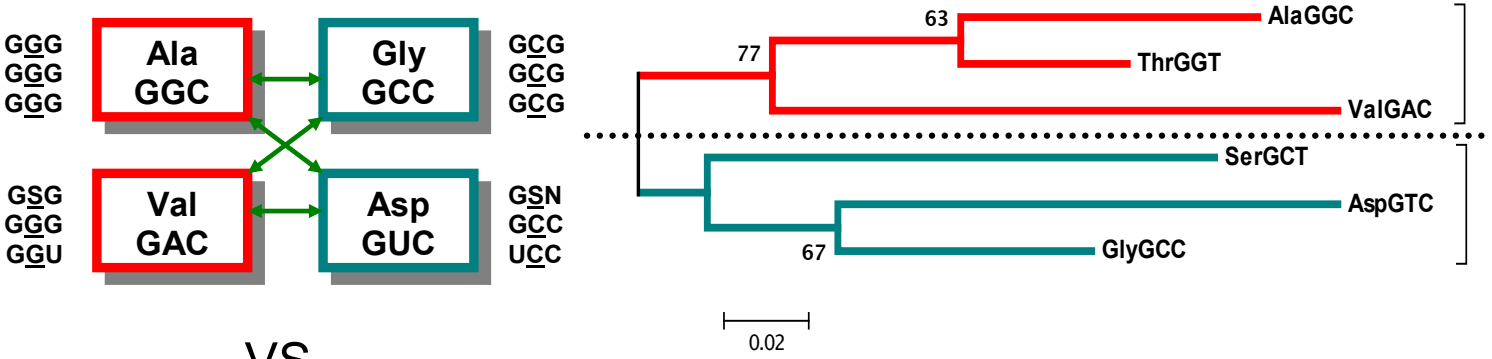


This is the second tetrad of amino acids that are complementary at the 2nd position of the acceptor stem in all combinations (legitimate W-C and "wobbling" pairs)

A

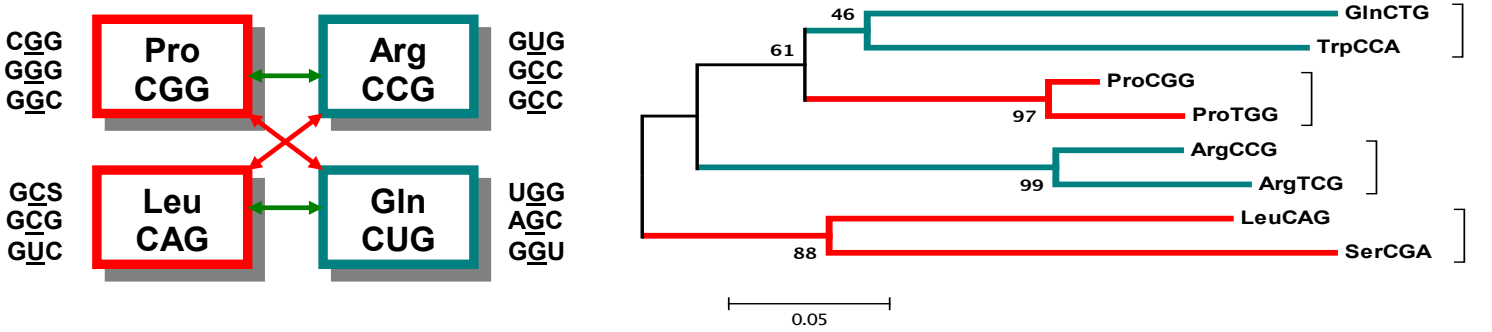

Gly Ala
5′ - GGC --- GCC ->3′
3′<- CCG --- CGG --5′
Ala Gly
Gly Ala
5′ - GGC --- GCC ->3′
5′ - GGU --- GUC ->3′
Gly Val
Gly Val
5′ - GGU --- GUC ->3′
3′<- CCA --- CAG - 5′
Thr Asp

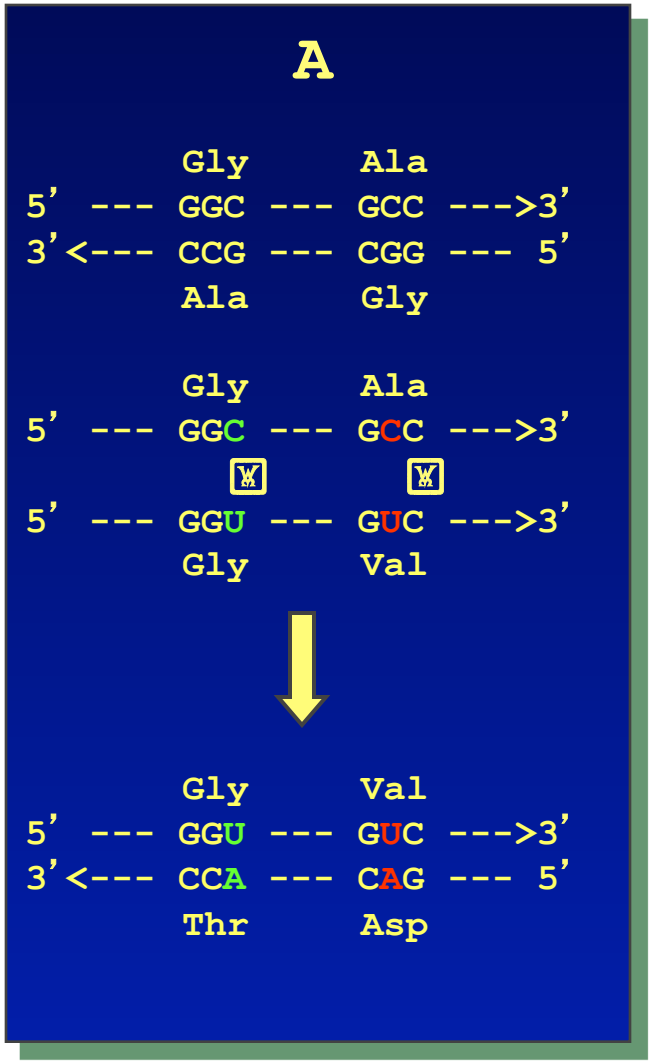
B

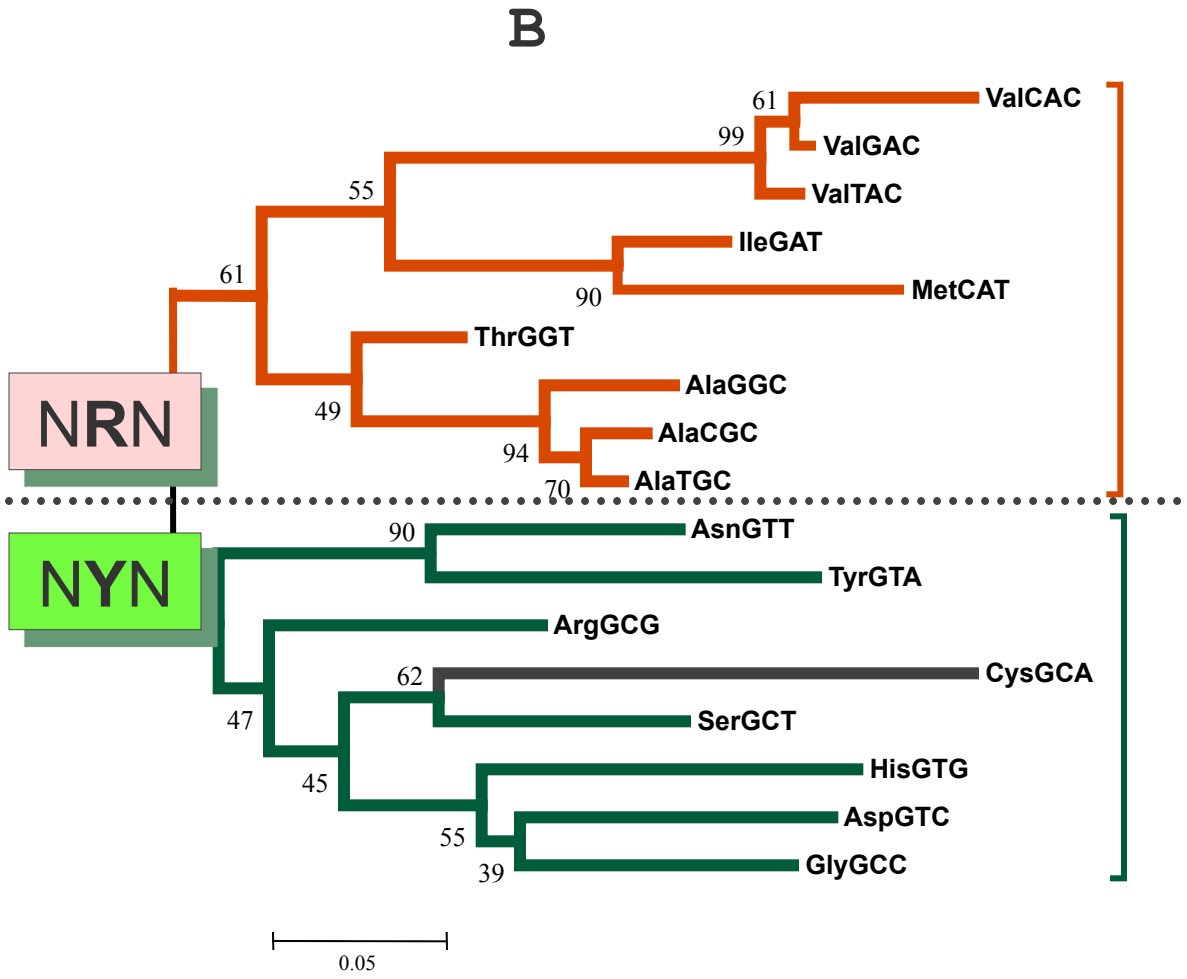

GGG GGG GGG
Ala GGC
Gly GCC
GCG GCG GCG
GSG GGG GGU
Val GAC
Asp GUC
GSN GCC UCC
63
77
AlaGGC
ThrGGT
ValGAC
SerGCT
AspGTC
67
GlyGCC
0.02

VS.

C

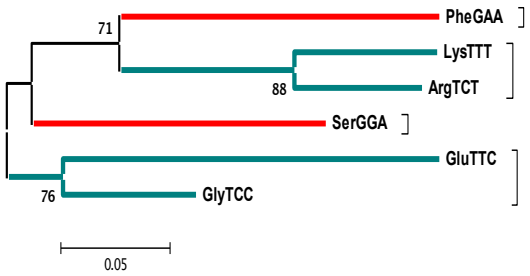

CGG GGG GGC
Pro CGG
Arg CCG
GUG GCC GCC
GCS GCG GUC
Leu CAG
Gln CUG
UGG AGC GGU
46
61
GlnCTG
TrpCCA
ProCGG
97
ProTGG
ArgCCG
99
ArgTCG
LeuCAG
88
SerCGA
0.05

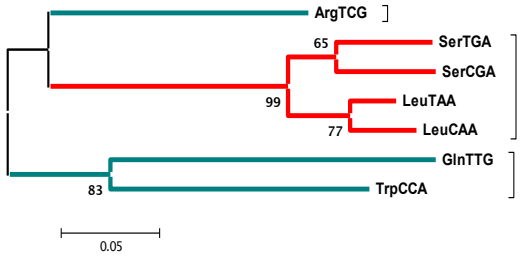


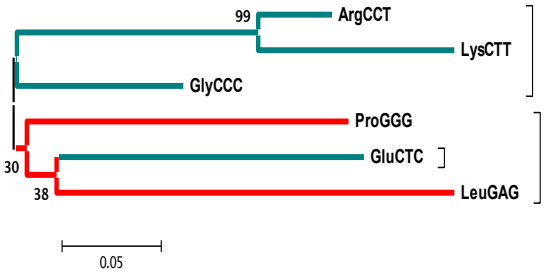


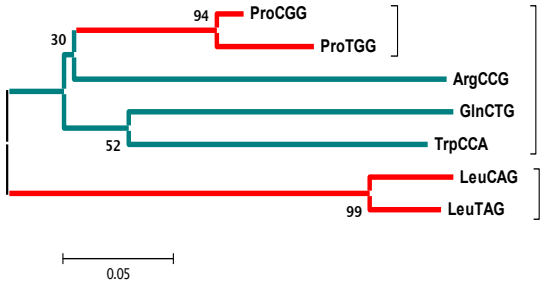
A: The scheme explaining how the double-strand coding could assist the genetic code to expand in pairs of complementary codons and corresponding amino acids. Shown are missense GCC→GUC (Ala→Val) and silent GGC→GGU (Gly→Gly) transitions in one strand complemented by GGC→GAC(Gly→Asp) and conservative GCC→ACC(Ala→Thr) transitions in the opposite strand.

B: The NJ-phylogenetic tree for ancestral tRNAs of archaebacteria generated by the two tetrads of amino acids (showing the dual complementarity in all four combinations of the 2nd base pairing in anticodons and symmetrically codons: not only G–C and A–U but also G–U and A–C) and their closest, one-transition-step-distanced, mutational derivatives. These tetrads are [Ala (GGC), Gly (GCC), Val (GAC), Asp (GAC)], and [Ala(CGC), Arg(GCG), Val(CAC), His(GUG)}. Note that this is the only tree that exhibits a distinct NRN vs. NYN bilateral pattern of branching, and that it covers 13 amino acids, 17 anticodons and 32 codons.


71
PheGAA
LysTTT
88
ArgTCT
SerGGA
GluTTC
76
GlyTCC
0.05

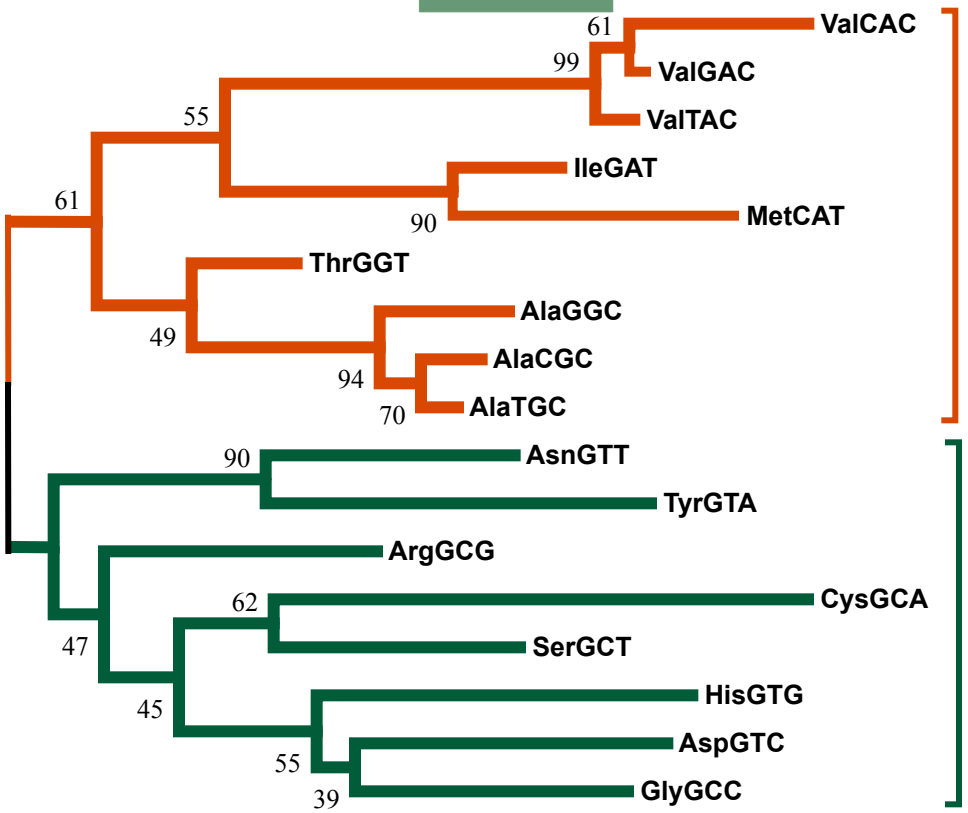

ArgTCG
65
SerTGA
SerCGA
99
LeuTAA
77
LeuCAA
GlnTTG
83
TrpCCA
0.05


99
ArgCCT
LysCTT
GlyCCC
ProGGG
30
GluCTC
38
LeuGAG
0.05

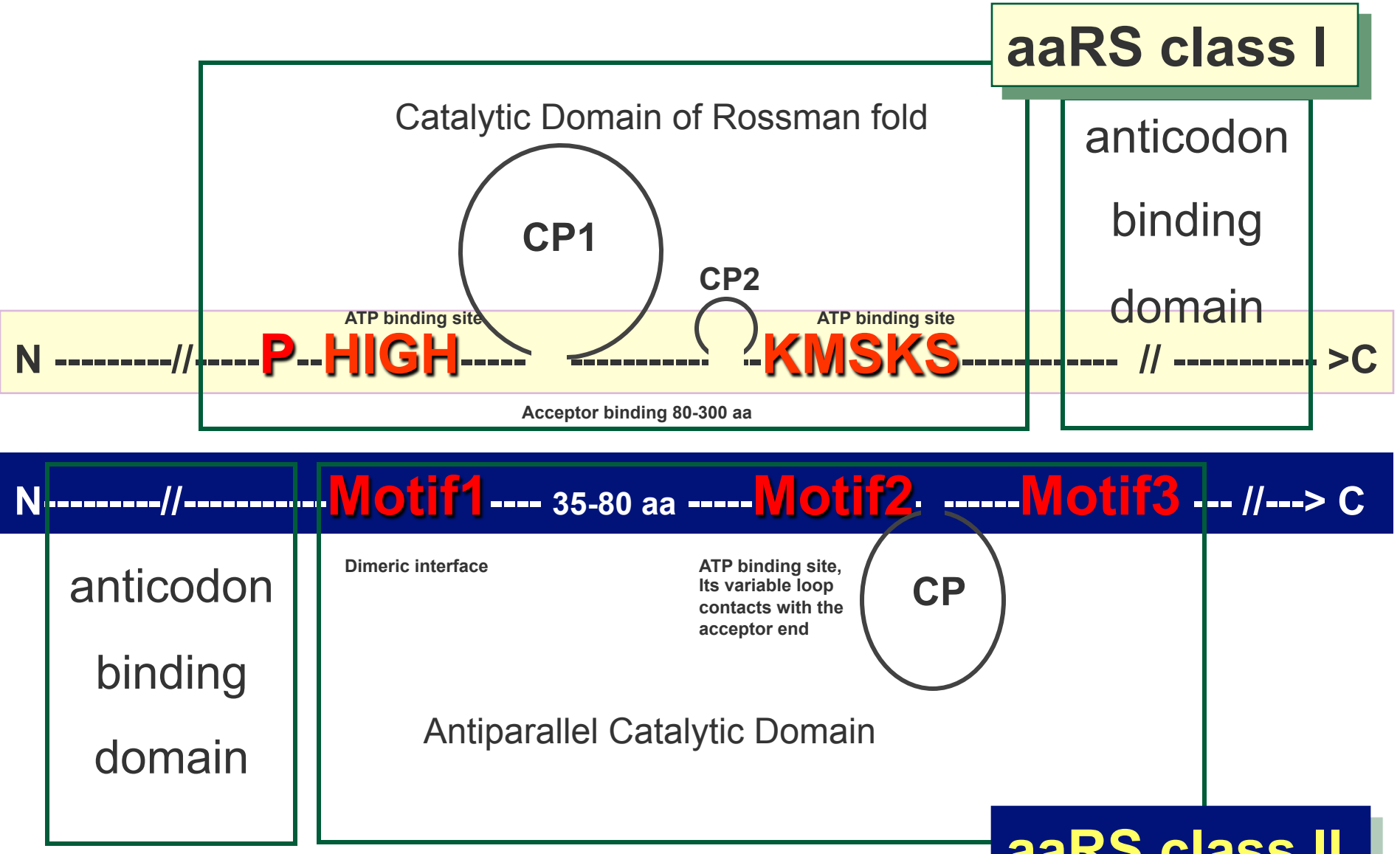

94
ProCGG
30
ProTGG
ArgCCG
GlnCTG
52
TrpCCA
LeuCAG
99
LeuTAG
0.05

VS.

NRN


61
ValCAC
99
ValGAC
ValTAC
55
IleGAT
61
90
MetCAT
ThrGGT
AlaGGC
49
AlaCGC
94
AlaTGC
70
90
AsnGTT
TyrGTA
ArgGCG
62
CysGCA
47
SerGCT
HisGTG
45
AspGTC
55
GlyGCC
39
0.05

NYN



Motif 1: + G ▣ xx ▣ xx P ▣ ▣

Motif 2: + ▣ ▣ +/- ▣ xxx FR x E/D… … + ▣ xx – F xxx - ▣ x ▣ ▣

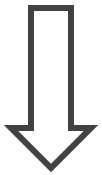
Motif 3: G ▣ G ▣ G ▣ E R ▣ ▣ ▣ ▣ ▣

From: (Rodin & Ohno, 1995)

At some turning point of the code evolution P-AARSs, as better catalysts, began to replace R-AARSs,

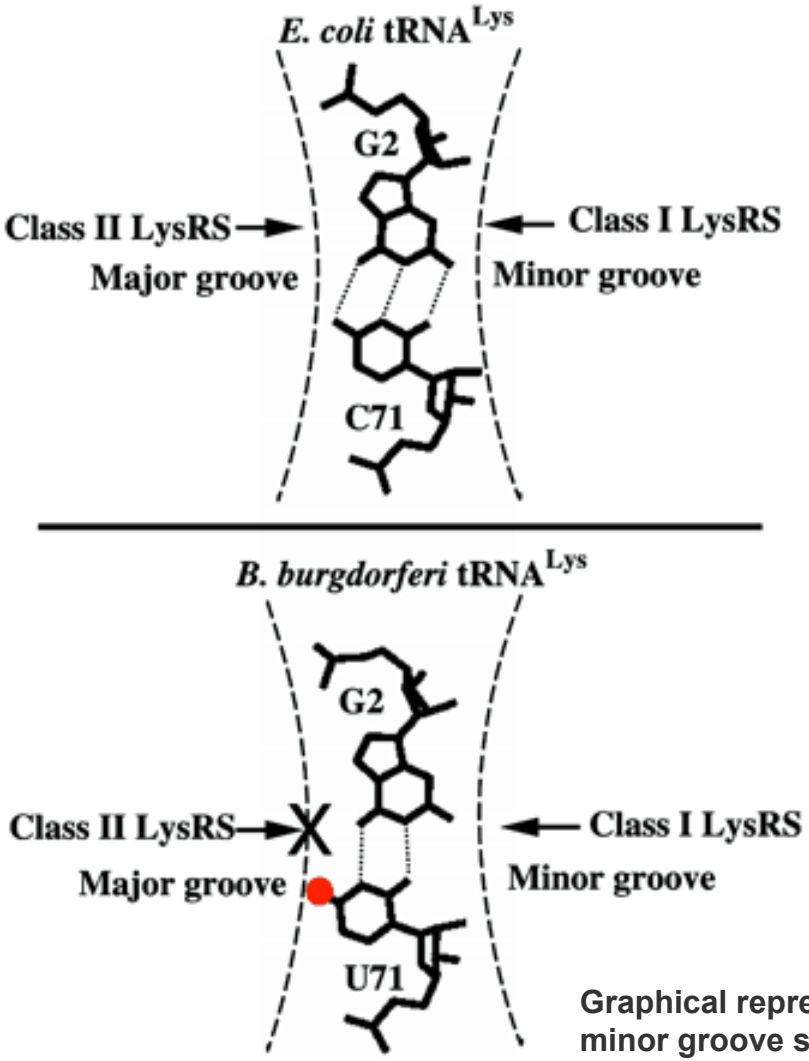
The principle of evolutionary continuity implies for these replacements to **maintain the adaptive features of ribozymic aminoacylation**

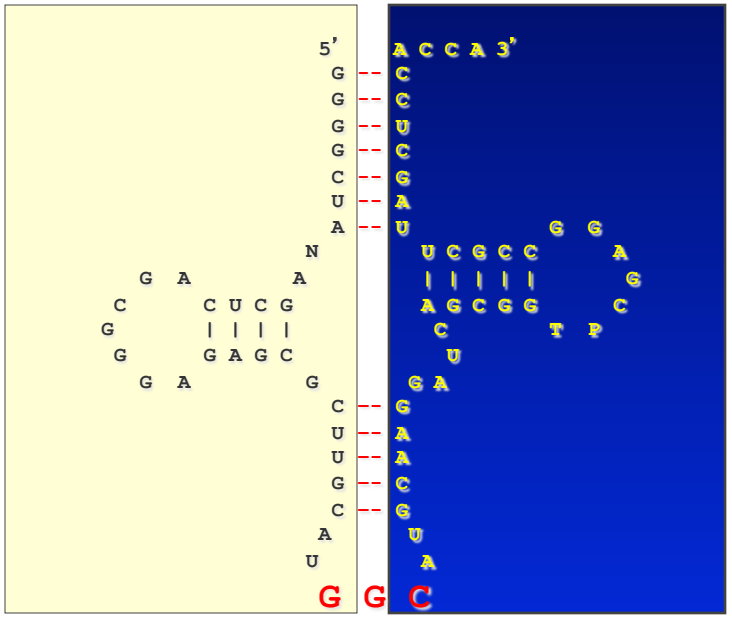
- translation of both complementary strands in first protein-encoding genes,
- the complementary triplets-based binary sub-code for aminoacylation at 2’-OH (minor groove side) and 3’-OH (major groove side) ends by the two putative R-AARSs,
- the complementarity of the two putative R-AARSs themselves
- their possible direct linkage with the genes for protein successors



Class I and II P-AARSs have also arisen from the complementary strands of one and the same ancestral gene

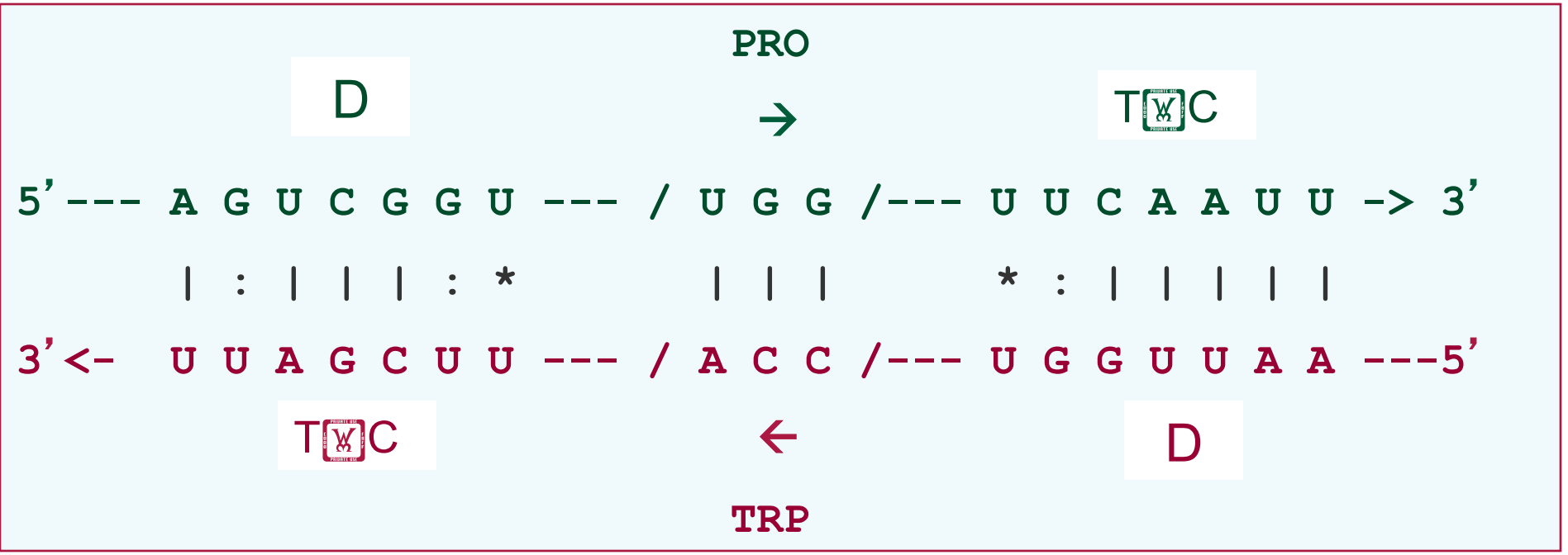
Co-evolution (OK!) vs. Co-revolution (?)...





Graphical representation of the atomic environments of the major and minor groove sides at the 2:71 position of *E. coli* tRNA^{Lys}. Adapted from Schimmel and Ribas de Pouplana (1999, 2004)

T  C vs. D "head-to-tail" comparison



THREE ORIGINS

CODE

Imprints of primordial complementary symmetry of the code in tRNAs and aaRSs

GENES

Flexible epigenetic mechanisms in evolution by gene duplications

CANCER

Origin and selection of carcinogenic mutations in quiescent stem cells

City of Hope®